Automated Morphometric Characterization of the Cerebral Cortex for the Developing and Ageing Brain

Manuel Jorge Machado Cardoso

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Centre for Medical Image Computing

University College London

2012

I, Manuel Jorge Machado Cardoso, confirm that the work presented in this thesis is my own.

Where information has been derived from other sources,

I confirm that this has been indicated in the thesis.

To Carolina, my parents and my sister Sara

Abstract

Morphometric characterisation of the cerebral cortex can provide information about patterns of brain development and ageing and may be relevant for diagnosis and estimation of the progression of diseases such as Alzheimer's, Huntington's, and schizophrenia. Therefore, understanding and describing the differences between populations in terms of structural volume, shape and thickness is of critical importance. Methodologically, due to data quality, presence of noise, PV effects, limited resolution and pathological variability, the automated, robust and time-consistent estimation of morphometric features is still an unsolved problem.

This thesis focuses on the development of tools for robust cross-sectional and longitudinal morphometric characterisation of the human cerebral cortex. It describes techniques for tissue segmentation, structural and morphometric characterisation, cross-sectional and longitudinally cortical thickness estimation from serial MRI images in both adults and neonates.

Two new probabilistic brain tissue segmentation techniques are introduced in order to accurately and robustly segment the brain of elderly and neonatal subjects, even in the presence of marked pathology. Two other algorithms based on the concept of multi-atlas segmentation propagation and fusion are also introduced in order to parcelate the brain into its multiple composing structures with the highest possible segmentation accuracy. Finally, we explore the use of the Khalimsky cubic complex framework for the extraction of topologically correct thickness measurements from probabilistic segmentations without explicit parametrisation of the edge. A longitudinal extension of this method is also proposed.

The work presented in this thesis has been extensively validated on elderly and neonatal data from several scanners, sequences and protocols. The proposed algorithms have also been successfully applied to breast and heart MRI, neck and colon CT and also to small animal imaging. All the algorithms presented in this thesis are available as part of the open-source package NiftySeg.

Acknowledgements

I would like to thank my primary supervisor, Prof. Sébastien Ourselin, for his advices and guidance over the last few years. I am also very grateful for his support and friendship.

My gratitude goes as well to my secondary supervisor, Prof. Nick Fox, who provided me valuable guidance and support over the course of my thesis.

I would like to thank all my CMIC (Centre for Medical Image Computing) colleagues and friends, specially Marc, Pankaj and Abi, who made our group an excellent place to work.

I am very thankful to everyone I collaborated with over these last years within the Centre for Medical Image Computing, the Dementia Research Centre, the Wellcome Trust Centre for Neuroimaging, the Department of Clinical and Experimental Epilepsy and the Centre for Advanced Biomedical Imaging.

A very special thanks to Carolina.

Publication List

Peer-reviewed Journal Papers

1. **Cardoso, M. J.**, Melbourne, A., Kendall, G.S., Modat, M., Robertson, N.J., Marlow, N., Ourselin, S., AdaPT: adaptive preterm segmentation algorithm for neonatal brain MRI. (2012) Neuroimage

2. **Cardoso, M. J.**, Modat, M., Cash, D., Leung, K. K., Barnes, J., Fox, N.C., Ourselin, S., The Alzheimer's Disease Neuroimaging Initiative, STEPS: Similarity and Truth Estimation for Propagated Segmentations. (2012) Medical Image Analysis (Under Submission)

3. Mertzanidou, M. J., Hipwell, J., **Cardoso, M. J.**, Zhang, X., Tanner, C., Ourselin, S., Bick, U., Huisman, H., Karssemeijer, N., Hawkes., D., MRI to X-ray mammography registration using a volume-preserving affine transformation (2012) Medical Image Analysis

4. Han, L., Hipwell, J., Tanner, C., Taylor, Z., Mertzanidou, T., **Cardoso, M. J**, Ourselin, S., Hawkes, D., Development of patient-specific biomechanical models for predicting large breast deformation (2012) Physics in Medicine and Biology

5. Daga, P, Winston, G, Modat, M, White, M, Mancini, L, **Cardoso, M. J**, Symms, M., Hawkes, D, Duncan, J, Ourselin, S., Accurate Localisation of Optic Radiation during Neurosurgery in an Interventional MRI Suite. (2011) IEEE Trans Med Imaging

6. Roth, H. R. , McClelland, J. R. , Boone, D. J., Modat, M., **Cardoso, M. J**., Hampshire, T. E., Hu, M., Ourselin ,S., Slabaugh, G. G., Halligan, S., Hawkes, D. J., Registration of the endoluminal surfaces of the colon derived from prone and supine CT colonography. (2011) Medical Physics.

7. Clarkson, M. J., **Cardoso, M. J.**, Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., Fox, N. C., Ourselin, S., A comparison of voxel and surface based cortical thickness estimation methods. (2011) Neuroimage

8. **Cardoso, M. J**., Clarkson, M. J., Ridgway, G. R., Modat, M., Fox, N. C., Ourselin, S., The Alzheimer's Disease Neuroimaging Initiative, LoAd: A locally adaptive cortical segmentation algorithm, (2011) Neuroimage

Peer-reviewed Conference Papers

1. **Cardoso, M. J.**, Wolz, R., Modat, M., Rueckert, D., Ourselin, S.: Geodesic Information Flow. (2012) MICCAI

2. **Cardoso, M. J.**, Winston, G., Modat, M., Keihaninejad, S., Duncan, J., Ourselin, S.: Geodesic Shape Based Average. (2012) MICCAI

3. **Cardoso, M. J.**, Modat, M., Ourselin, S., Multi-label similarity and truth estimation for propagated segmentations (STEPS) validation (2012) MICCAI MASLF

4. Pedemonte, S., **Cardoso, M. J.**, Arridge, S., Hutton, B.F., Ourselin, S.; Steady-state model of the radio-pharmaceutical uptake for MR-PET (2012) MICCAI

5. Cash, D., Melbourne A., Modat, M., **Cardoso, M.J.**, Clarkson, M.J., Fox, N., Ourselin, S., Cortical Folding Analysis on Patients with Alzheimer's Disease and Mild Cognitive Impairment (2012) MICCAI

6. Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Ourselin, S., Radial structure in the preterm cortex; persistence of the preterm phenotype at term equivalent age? (2012) MICCAI

7. Ma, D., **Cardoso, M. J.**, Modat, M., Ourselin, S., Multi Atlas Segmentation applied to in vivo mouse brain MRI, (2012) MICCAI MASLF

8. Modat, M., Leung, K., **Cardoso, M. J.**, Ourselin, S., Atrophy measurement based on segmentation propagation and the boundary shift integral technique (2012) MICCAI Atrophy

9. Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Ourselin, S., A cortical surface analysis of very preterm infants on term-equivalent age MRI (2012) MICCAI PAPI

10. Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Ourselin, S., NeoBrainS12 Challenge: Adaptive neonatal MRI brain segmentation with myelinated white matter class and automated extraction of ventricles I-IV (2012) MICCAI NeoBrainS

11. Zuluaga, M., **Cardoso, M. J.**, Ourselin, S., Automatic Right Ventricle Segmentation using Multi-Label Fusion in Cardiac MRI (2012) MICCAI RVSC

12. Modat, M., **Cardoso, M. J.**, Daga, P., Cash, D., Fox, N., Ourselin S.: Inverse-Consistent Symmetric Free Form Deformation (2012) Workshop on Biomedical Image Registration, p79-88

13. Young, J., Modat, M., **Cardoso, M. J.**, Ashburner, J., Ourselin, S., Classification of Alzheimer's disease patients and controls with Gaussian processes (2012) ISBI

14. **Cardoso, M. J.**, Modat, M., Keihaninejad, S., Cash, D., Ourselin, S.: Multi-STEPS: Multi-label Similarity and Truth Estimation for Propagated Segmentations. (2012) MMBIA

15. Modat, M., Ridgway, G., Daga, P., **Cardoso, M. J.**, Ashburner, J., Ourselin, S., Parametric non-rigid registration using a stationary velocity field, (2012), MMBIA.

16. Daga, P., Winston, G., Modat, M., **Cardoso, M. J.**, White, M., McEvoy, A. W., Thornton, J., Hawkes, D., Duncan, J., Ourselin, S., Improved neuronavigation through integration of intra-operative anatomical and diffusion images in an interventional MRI suite, (2011), IPCAI.

17. Modat, M., Ridgway, G. R., Daga, P., **Cardoso, M. J.**, Hawkes, D. J., Ashburner, J., Ourselin, S., Log-Euclidean free-form deformation, (2011), SPIE.

18. Clarkson, M. J., **Cardoso, M. J.**, Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., Fox, N. C., Ourselin, S.,: A Comparative Study of Voxel and Surface Based Cortical Thickness Methods in Frontotemporal Dementia , (2011), AAIC

19. **Cardoso, M. J**., Clarkson, M.J., Modat, M., Ourselin, S.: Longitudinal Cortical Thickness Estimation using Khalimsky's Cubic Complex. (2011) MICCAI

20. **Cardoso, M. J**., Melbourne A., Kendall G. S., Modat M, Hagmann C. F., Robertson N. J., Marlow N., Ourselin S., Adaptive Neonate Brain Segmentation. (2011) MICCAI

21. **Cardoso, M. J.** Leung, K. K., Modat, M., Barnes, J., Ourselin, S.: Locally Ranked STAPLE for template based segmentation propagation. (2011) MICCAI - MALSF

22. Daga, P., Winston, G., Modat, M., **Cardoso, M. J.**, Stretton, J., Symms, M., McEvoy, A. W., Duncan, J., Ourselin, S., Integrating structural and diffusion mr information for optic radiation localisation in focal epilepsy patients, (2011), IEEE ISBI.

23. Clarkson, M. J., **Cardoso, M. J.**, Modat, M., Leung, K. K., Rohrer, J., Fox, N. C., Ourselin, S. Cross-sectional analysis using voxel or surface based cortical thickness methods: a comparison study, (2011), IEEE ISBI.

24. **Cardoso, M. J**., Clarkson, M. J., Modat, M., Ourselin, S., On the Extraction of Topologically Correct Thickness Measurements using Khalimsky's Cubic Complex, (2011), IPMI

25. Daga, P., Modat, M., Micallef, C., Mancini, L., White, M., **Cardoso, M. J.**, Kitchen, N., McEvoy, A. W., Thornton, J., Yousry, T., Hawkes, D., Ourselin, S., Near real time brain shift estimation for interventional MRI suite, (2010), HP-MICCAI

26. Pedemonte, S., **Cardoso, M. J.**, Bousse, A., Panagiotou, C., Kazantsev, D., Arridge, S., Hutton, B.F., Ourselin, S.; Class conditional entropic prior for MRI enhanced SPECT reconstruction (2010) IEEE NSS/MIC

27. Mertzanidou, T., Hipwell, J. H., **Cardoso, M. J.**, Tanner, C., Ourselin, S., Hawkes, D. J., X-ray Mammography - MRI Registration Using a Volume-Preserving Affine Transformation and an EM-MRF for Breast Tissue Classification, (2010), IWDM.

28. **Cardoso, M. J**., Clarkson, M. J. , Ridgway, G. R., Modat, M. , Talbot, H., Couprie, M., Ourselin, S., Topologically correct cortical segmentation using Khalimsky's cubic complex framework, (2010), SPIE

29. **Cardoso, M. J.**, Clarkson, M. J., Modat, M., Ridgway, G. R., Ourselin, S. Locally weighted Markov random fields for cortical segmentation, (2010), IEEE ISBI

30. **Cardoso, M. J.**, Clarkson, M. J., Modat, M., Ridgway, G. R., Fox, N.C., Ourselin, S: Improved Maximum a Posteriori Cortical Segmentation by Iterative Relaxation of Priors. (2009) MICCAI

Peer-reviewed Conference Abstracts

1. Keihaninejad, S., Zhang, H., Shakespeare, T., Ryan, N., Malone, I., Frost, C., **Cardoso, M. J.**, Modat, M., Lehmann, M., Warren, J., Cash, D., Crutch, S., Ourselin, S., Fox, N.; Longitudinal changes in fiber tract integrity in posterior cortical atrophy: Serial diffusion tensor imaging (2012) Alzheimer's & Dementia 8 (4), p75

2. Keihaninejad, S., Zhang, H., Shakespeare, T., Ryan, N., Malone, I., Frost, C., **Cardoso, M. J.**, Modat, M., Lehmann, M., Warren, J., Cash, D., Crutch, S., Ourselin, S., Fox, N.; White matter damage in posterior cortical atrophy assessed in vivo using diffusion tensor magnetic resonance imaging (2012) Alzheimer's & Dementia 8 (4), p176-177

3. Keihaninejad, S., Zhang, H., Shakespeare, T., Ryan, N., Malone, I., Frost, C., **Cardoso, M. J.**, Modat, M., Lehmann, M., Warren, J., Cash, D., Crutch, S., Ourselin, S., Fox, N.; White matter tract changes in mild-to-moderate Alzheimer's disease revealed by tensor-based registration analyses of diffusion tensor imaging (2012) Alzheimer's & Dementia 8 (4), p31-32

4. Keihaninejad, S., Zhang, H., Shakespeare, T., Ryan, N., Malone, I., Frost, C., **Cardoso, M. J.**, Modat, M., Lehmann, M., Warren, J., Cash, D., Crutch, S., Ourselin, S., Fox, N.; Longitudinal changes in fiber tract integrity in Alzheimer's disease: A serial diffusion tensor imaging study (2012) Alzheimer's & Dementia 8 (4), p30-31

5. Kinnunen, K, Cash, D., Leung, K., Liang, K., **Cardoso, M. J.**, Malone, I., Frost, C., Ourselin, S., Fox, N.; Brain and hippocampal rates of atrophy in familial Alzheimer's disease mutation carriers: Preliminary findings from the DIAN study (2012) Alzheimer's & Dementia 8 (4), p72-73

6. Kinnunen, K, Cash, D., Leung, K., Liang, K., **Cardoso, M. J.**, Malone, I., Frost, C., Ourselin, S., Fox, N.; Cross-sectional cerebral volumetric differences and associations with estimated time to age-at-onset in familial Alzheimer's disease: Findings from the DIAN study (2012) Alzheimer's & Dementia 8 (4), p72-73

7. Andrews, K., Modat, M., Macdonald, K., Yeatman, T., **Cardoso, M. J.**, Leung, K., Barnes, J., Fox, N., Ourselin, S., Schott, J., Cross-sectional cerebral volumetric differences and associations with estimated time to age-at-onset in familial Alzheimer's disease: Findings from the DIAN study (2012) Alzheimer's & Dementia 8 (4), p22-p23

8. Cash, D., Liang, Y., Leung, K., Ryan, N.,**Cardoso, M. J.**, Barnes, J., Fox, N., Ourselin, S.; Rates of brain and hippocampal atrophy in presymptomatic familial Alzheimer's disease: Acceleration and mutation effects (2012) Alzheimer's & Dementia 8 (4), p18-p19

9. Melbourne A., Kendall, G.S., **Cardoso, M.J.**, Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Analysing the cortical folding pattern of very preterm neonates scanned at term-equivalent age: Correlations with diffusion tensor tractography (2012) ISMRM

10. Melbourne A., Kendall, G.S., **Cardoso, M.J.**, Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Analysing the cortical folding pattern of very preterm neonates scanned at term-equivalent age: Correlations with diffusion tensor tractography (2012) ISMRM

11. **Cardoso, M.J.**, Melbourne A., Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Outlier Rejection for Adaptive Neonatal Segmentation (2012) ISMRM

12. **Cardoso, M.J.**, Melbourne A., Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Adaptive neonatal brain segmentation: application to ventriculomegaly and excess extra-axial cerebral-spinal fluid. (2011) ESPR

13. Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Automated Analysis of the Preterm Neonatal Cortex at Term Equivalent Age and Correlation with Cognitive Outcome at 1 Year Corrected Age (2011) ESPR

14. Brazdova, V., **Cardoso, M.J.**, Chupin, M, Ourselin S., Lemieux L., Computational model of brain atrophy (2011) OHBM.

Contents

List of Figures

List of Tables

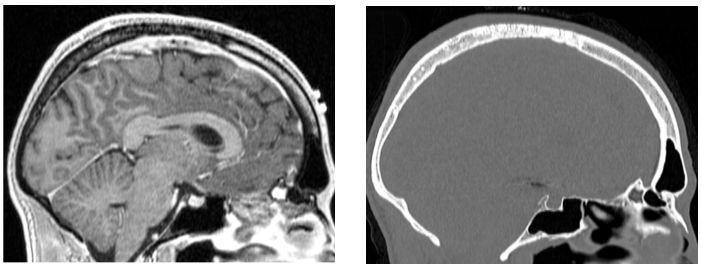
Chapter 1

Introduction

Imaging is currently an ubiquitous component of clinical procedure. Medical images, acquired using many different techniques - ultrasound (US), computed tomography (CT), positron emission tomography (PET), single-photon emission computed tomography (SPECT), magnetic resonance imaging (MRI) and its different weightings (T1, T2, PD), etc - provide rich and complementary information about the underlying anatomy, physiology and tissue microstructure. The aim of medical image analysis techniques is to explore this rich source of information, to study human anatomy and physiology and to detect, diagnose and examine disease processes.

Anatomical imaging techniques, like CT and MRI, provide information about certain properties of the anatomy under observation. If different tissues have different physical or chemical properties, the signal obtained from these tissues will differ, leading to the existence of contrast between tissues. As an example, Fig 1.1 shows two aligned sagittal slices of a human brain from two different modalities (MRI and CT), each one presenting different tissue contrasts. The contrast between different tissues can thus be exploited to localise and study meaningful anatomical structures.

In the realm of medical imaging and human anatomy, quantitative morphology, normally referred to as morphometry, encompasses the study of size (volume, thickness) and shape of one or several structures or organs of interest. As most pathological processes directly affect the morphology of specific structures, studying how the morphometry varies between a normal and a diseased population is of crucial importance. These morphological differences can lead to the development of imaging biomarkers - an image-based indicator of a biological physiological and/or pathological state. These imaging biomarkers can thus enable a more quantitative diagnosis and estimation of disease progression.



a *b*

Figure 1.1: A sagittal view of a human brain acquired using (a) T1-weighted MRI and (b) CT

As an example, clinically important patterns of longitudinal brain change in pathological and normal brains, either during development or ageing, may be detected through the analysis of serial MRI. It has been shown that changes in volume of key anatomical structures (*e.g* hippocampus) and cortical thickness over time have an important correlation with normal development and ageing and also with various diseases such as Alzheimer's, Huntington's and schizophrenia (Jernigan and Tallal, 1990; Battin et al., 1998; Lerch et al., 2005; Lehmann et al., 2009; Rosas et al., 2008; Thambisetty et al., 2010). Thus, the extraction of structural volumes and cortical thickness measurements from brain images may be relevant for diagnosis and estimation of disease progression (Desikan et al., 2009), and can potentially be used in the monitoring of disease-modifying treatments.

This thesis will focus on the examples described above. I aim at developing new and robust techniques for image segmentation, parcellation and longitudinally consistent structural volumetry and cortical thickness estimation from serial MRI images. This chapter will present some of the motivation and current challenges in image segmentation, parcellation and cortical thickness estimation.

1.1 Tissue segmentation in brain MRI

Image segmentation, and more specifically brain tissue segmentation, is one of the most important preprocessing step in neuroimage analysis. This class of methods aim at simplifying the representation of an image, by locating certain anatomically relevant tissues, into something more meaningful and thus easier to analyse. More precisely, image segmentation is the process of assigning a label (or its probabilistic equivalent) to every voxel in an image such that voxel with the same label share certain physical, anatomical and/or functional characteristics. The example in Fig.1.2 shows a brain segmented in its three main composing tissues: white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF), where the brightness of the pixel in each image represents the probability to belong to that particular tissue.

The assumption behind most segmentation algorithms is that voxels from the same tissue will have a similar observed intensity, as the properties within the tissue should be approximately constant. However, as with most image acquisition systems, several artefacts can corrupt the image and subsequently this assumption. The most common artefact is the presence of noise, which is Rician distributed in MRI images. This effect can be seen in Fig. 1.3(b). The high amount of noise, and the low signal-to-noise ratio (SNR), can greatly hinder the quality of the segmentation, as the method needs to use contextual information to be able to correctly label certain tissues of interest. Another source of artefacts is MRI intensity non-uniformity (INU), caused by magnetic field inhomogeneity. This effect leads to a spatially smooth corruption of the image intensities, as seen in Fig. 1.3(c), where the same tissue can produce very different observed intensities. For example, the two areas highlighted in Fig. 1.3(c) should have very similar physical properties, but the observed intensities are significantly different. Even though noise and INU are the most common artefacts present in MRI images, several other imaging artefacts (*e.g.* magnetic susceptibility, chemical shift, patient movement, geometrical distortions) and limitations in terms of spatial resolution can also hamper the accuracy and stability of the segmentation procedure.

While imaging artefacts are still problematic, recent advances in acquisition hardware (*e.g.* mag-

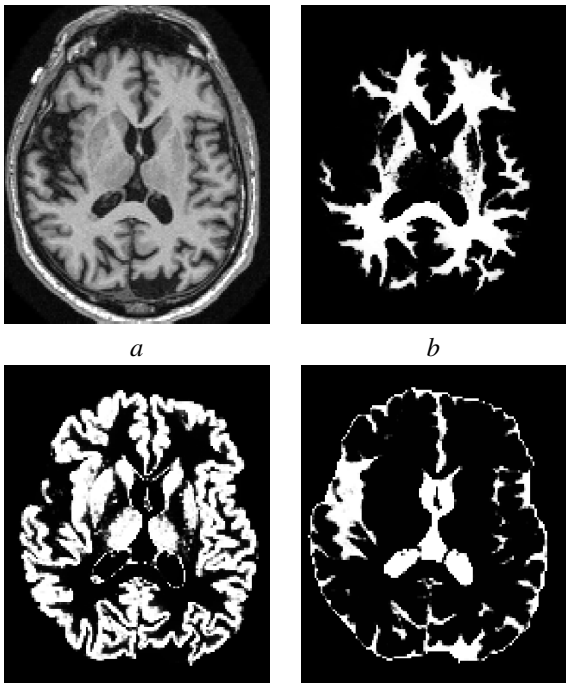


Figure 1.2: (a) T1-weighted MRI image of a human brain segmented in to its 3 main tissue classes (b) white matter, (c) grey matter and (d) cerebrospinal fluid.

netic field strength, receiver coils) and MRI sequences has greatly reduced their repercussions in the segmentation quality. However, even under tightly controlled acquisition protocols (as in clinical trials), accurately and robustly segmenting data from multiple sites and scanners without introducing bias still poses a challenge. Despite the above described challenges and limitations, much of the interest of brain segmentation has shifted from image acquisition problems to being able to deal with anatomically disparate, morphologically complex and pathological brains. More specifically, several new challenges arise from pathological cases: the presence of pathological elements in the image (*e.g.* tumours, cysts) can greatly increase the complexity of the segmentation procedure, as their spatial location and signal characteristics can vary between affected subjects in a population; the loss of contrast between WM and GM due to neurodegeneration and normal ageing can result in fuzzy tissue boundaries; the presence of

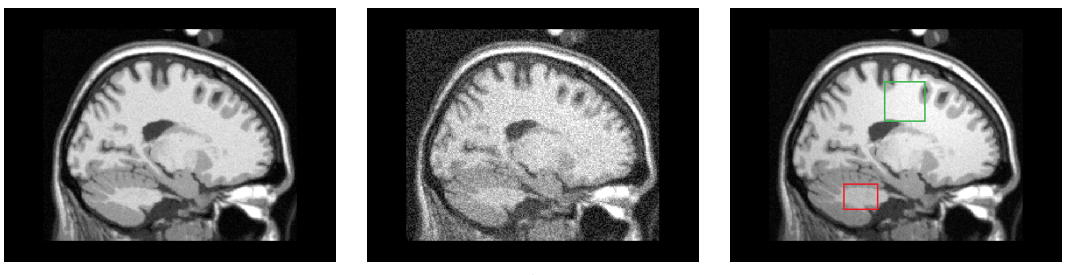


Figure 1.3: Three sagittal views of simulated T1-weighted MRI images of a human brain from the BrainWeb database: image with (a) low-noise and low INU, (b) high noise and (c) high INU. The red and green squares represent two areas from the same tissue with completely different intensities due to INU effects.

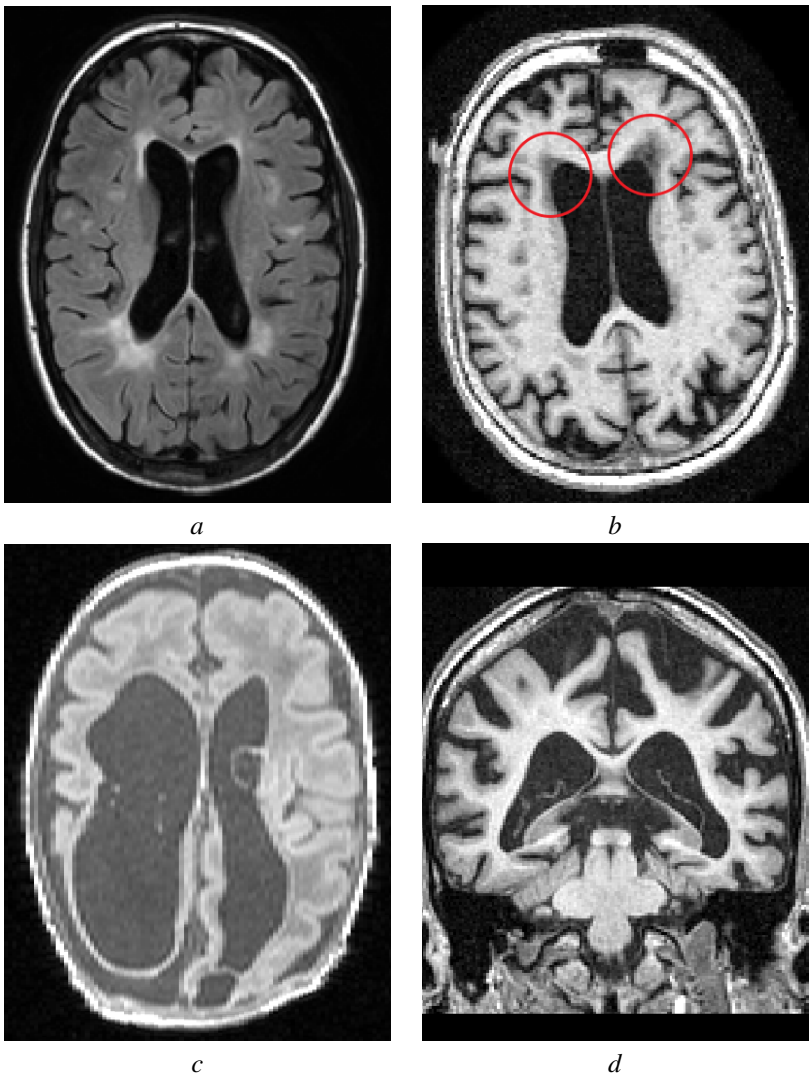
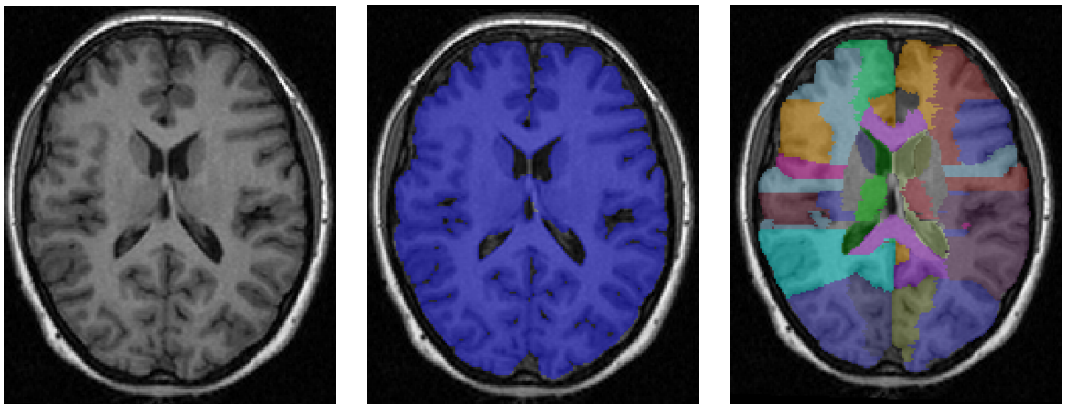


Figure 1.4: Four examples of challenging brains to segment: (a) FLAIR MRI of a multiple sclerosis patient with WM hiperintensities; (b) T1-weighted MRI image a dementia patient with periventricular WM dammage; (c) T1-weighted MRI image of a neonatal brain with ventriculomegaly, resulting in a colapsed cerebrum; (d) T1-weighted MRI image of an highly atrophiated brain of a patient with Alzheimer's disease.

outliers (*i.e.* unexpected elements in the image like WM lesions and micro-bleeds) can add an extra degree of complexity in the segmentation; and the presence of extreme morphologies (*e.g.* polymicrogiri, ventriculomegaly, severe atrophy) can also be very challenging. In foetal and neonatal brains, some non-pathological cases can also be challenging due to the neuro-developmental process and myelination. As some structures are still developing and myelinating, foetal and neonatal brains have different anatomical presentation and tissue properties at different time points. Some examples of these complex morphological and pathological presentations are shown in Fig. 1.4. Algorithmic solutions to achieve robust and unbiased segmentations of these complex cases will be explored throughout this thesis.



a *b* *c*

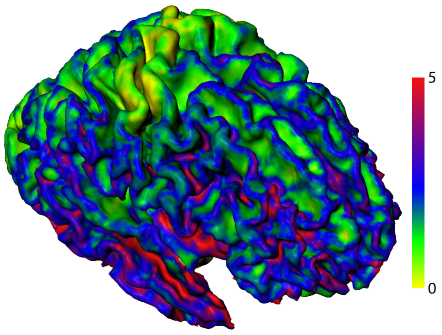
Figure 1.5: An example of structural parcellation: (a) T1-weighted MRI image of a human brain, (b) brain region-of-interest localization, (c) anatomical parcellation of the brain in its composing sub-structures according to the Hammers protocol (Hammers et al., 2003).

1.2 Structural parcellation in brain MRI

Structural parcellation is a key pre-processing step for selecting regions-of-interest (ROI), for morphometric analysis of anatomical structures (volume and shape) and for the estimation of pathological change and neurodevelopment. Although they are normally obtained from anatomical images (T1-weighted MRI), structural parcellations can be propagated to functional and quantitative imaging (FMRI, DWI, PET) modalities through the use of multi-modal image alignment.

In this context, structural parcellation refers to the process of labelling one or several anatomically or functionally distinct structures that can share similar signal properties, neighbouring spatial locations and be composed of more than one tissue type. The most common example of this problem is brain extraction, a process that aims at localising the boundary that separates brain from non-brain areas. In T1-weighted MRI, intensity alone does not allow us to separate these two ROIs without using contextual, shape and spatial constraints. Brain extraction is commonly used to simplify image processing pipelines, mainly tissue segmentation, as it limits the processing problem to a much more simplified anatomical region. The brain can also be separated into one or all of its composing anatomical sub-structures. For example, the amygdala and the hippocampus are two neighbouring structures involved in neurodegenerative pathologies. As these structures have similar image appearance, the process of parcellation also has to rely on geometrical and spatial constraints. Thus, instead of localising and parcelating each anatomical sub-structure independently, which can result in non-optimal and overlapping parcellations between neighbouring sub-structures, methods should focus on the simultaneous and unified parcellation of multiple brain structures. An example of brain extraction and parcellation of multiple anatomical sub-structures is shown in Fig. 1.5.

In brain imaging, image parcellation accuracy is mostly influenced by the presence of pathological and/or extreme morphologies and the limited number and quality of the training examples. The scarcity in the training data can result in limited knowledge about contextual and neighbouring rules in complex morphological cases, leading to non-optimal parcellations. Furthermore, the quality of the training data, commonly obtained by manual segmentation, and the variability between defined protocols can also



a *b*

Figure 1.6: The thickness measurements (side bar represents the thickness in millimetres) from one single patient using the KaTE method described in section 5.1.

limit the accuracy of the parcellation due to the introduction of errors and instability in the learned rules. Solutions for dealing with complex morphologies and with the limited amount and quality of training data are still very active areas of research that will be explored in this thesis.

1.3 Cortical Thickness

The automatic extraction of thickness measurements from anatomical structures has the potential to provide a biomarker for diagnosis and disease progression. More specifically, cortical thickness has been shown to be correlated with normal development (Jernigan and Tallal, 1990; Battin et al., 1998; Ajayi-Obe et al., 2000) and ageing (Shefer, 1973; Salat et al., 2004; Thambisetty et al., 2010) and also with various diseases such as Alzheimer's (Lerch et al., 2005; Du et al., 2007; Lehmann et al., 2009), Huntington's (Rosas et al., 2008), schizophrenia (Nesvåg et al., 2008). Notwithstanding its generalised usage, the reliable extraction of sub-voxel accurate measurements of thickness from probabilistic segmentations is still an unsolved problem.

Cortical thickness estimation (CTE) methods can be generally separated into surface-based and voxel-based techniques, each with its own advantages and disadvantages. Surface based methods typically fit a triangulated mesh to the region of interest whilst guaranteeing certain topological properties. These topological properties enable inter-patient comparison in a geometrically simplified conformally mapped spherical space (Fischl and Dale, 2000; Kim et al., 2005), increasing statistical power. However, as the parametrisation of the surface can be complex, the choice of the smoothness parameters and curvature constraints can bias the thickness measurements (Scott et al., 2009). On the other methodological side, voxel based methods use heuristic methods (line integrals, maximal sphere fitting, Laplacian integration) on the voxel grid to estimate cortical thickness, making them computationally very fast. As voxel-based methods work in the original discretised image space and not with surfaces, curvature constraint derived bias is not existent. However, the lack of topological guarantees in voxel-based methods does not allow inter-patient comparison on a conformally mapped spherical space. Thus, a method that combines the best of both methodologies (*i.e.* enforcing spherical topology without smoothness/curvature bias) would be ideal.

1.4 Thesis contributions

This thesis aims at developing robust and unbiased tools for tissue segmentation, structural parcellation and cortical thickness estimation and provide a comprehensive, fast and open-source set of algorithms to the neuroimage analysis community. Even though the thesis will focus on the analysis of neonatal and elderly human brains, the algorithms presented in this thesis are general enough to be applied to other organs, modalities, and even species. For example, some of the algorithms already have been successfully applied to breast segmentation and modelling (Mertzanidou et al., 2012), topology correction of endoluminal surfaces in colon CT (Roth et al., 2011), automated mesh generation for bio-mechanical simulations (Han et al., 2012) and more recently to small animal imaging, more specifically for mouse brain segmentation.

The contributions of this thesis include, but are not limited to the following:

- A probabilistic framework for brain segmentation developed specifically for elderly subjects that includes INU correction, spatial dependence via a Markov Random Field (MRF) and explicit modelling of partial-volume (PV) containing voxels and three novel modifications introduced to reduce the influence of the priors in an anatomically coherent way and improve the PV estimation and the delineation of deep sulci and gyri.

- A neonatal brain segmentation method that allows, for the first time, the segmentation of severely pathological neonatal subjects (e.g ventriculomegaly, cysts, sub-arachnoid CSF). This method adapts some of the ideas from the segmentation of elderly subjects to neonatal cases by generating patient specific atlases and by constraining the space of solutions of the segmentation model's parameters.

- A new image parcellation method based on the STAPLE label fusion algorithm (Warfield et al., 2004). This method incorporates a local similarity metric which estimates the expected image-based performance of each classifier in a voxel-by-voxel basis and also introduces a new MRF model optimised iteratively with a mean-field approximation in order to add spatial consistency and smoothness between the best local classifiers. This technique provides almost human-level accuracy for hippocampus segmentation and brain extraction and is currently being used as part of the Dementia Research Centre processing pipeline for image analysis and clinical trials.

- A new technique that improves parcellation performance when the subject to be segmented is morphologically very different from the subjects in the template database. This technique enables the propagation of anatomical parcellations, brain masks and even anatomical priors to morphologically disparate subjects by leaning the spatially varying embedding (*i.e.* the local manifold structure) of the data. A provisional patent on the method has been filed.

- A new voxel-based and topologically correct thickness estimation algorithm using the Khalimsky cubic complex framework. In this technique, the topology is implicitly corrected within the algorithm and cortical thickness is estimated as the integral of the geodesic path that crosses the cortex.

- A new way to enforce longitudinal consistency in cortical thickness measurements without directly constraining the thickness measurements. This new technique makes use of time-consistent transformations and Laplacian correspondence maps, resulting in more robust measurements of cortical thickness and significative group differences.

1.5 Thesis organisation

The next chapter will present a short review of the current state-of-the-art, advantages and limitations of the of the three main subjects for this thesis: segmentation, parcellation and cortical thickness estimation. Chapter 3 describes the contributions of this thesis regarding brain tissue segmentation in both elderly patients and neonatal brains, *i.e.* the LoAd (**Lo**cally **Ad**aptive segmentation) and the AdaPT (**Ada**ptive **P**re-**T**erm segmentation) algorithms. I show that the proposed methods are advantageous in terms of segmentation robustness and accuracy, segmentation bias and direct application to cortical thickness estimation, when compared to state-of-the-art tissue segmentation algorithms. In Chapter 4, I present my contributions to brain extraction and structural parcellation, both important for accurate tissue segmentation and cortical thickness estimation. The algorithms developed in this section are shown to be very accurate when compared to a gold-standard manual segmentation, even in the presence of pathological or morphologically complex subjects. Due to their outstanding performance and robustness, these parcellation algorithms are currently being used as the *de facto* processing pipeline for image analysis and clinical trials at the Dementia Research Centre.

As the successful development of highly accurate algorithms for tissue segmentation and brain extraction/parcellation fulfilled all the pre-requirements for cortical thickness estimation, Chapter 5 focuses of the development of a new set of tools to explore both cross-sectional and longitudinal cortical thickness estimation measurements. Lastly, Chapter 6 details both my open-source/open-science software effort and the software package (NiftySeg) that contains all the algorithms developed during my PhD and Chapter 7 concludes the thesis and outlines some future research directions.

Chapter 2

State-of-the-art

2.1 Tissue Segmentation

The process of tissue segmentation aims at simplifying the representation of an image by localising and labelling specific regions of interest, which share certain visual, anatomical or functional characteristics. Segmentation techniques can be divided into classes in many different ways, depending on the classification scheme:

1. Manual, semi-automatic and automatic
2. Pixel- and region-based
3. Model-based (*e.g.* probabilistic, level sets, active contours) and non model-based segmentation (*e.g.* thresholding, region growing, watershed)

On its simplest form, a tissue segmentation can be obtained by selecting a range of brightness values from a set of observed image intensities, followed by assigning the pixels that fall within or outside this range to the foreground and background classes respectively. This selection process, usually called thresholding, can be set interactively by a user or automatically through some particular algorithm. Automated thresholding algorithms separate the foreground and background classes using empirical (*e.g.* minimising intra-class variance, nearest neighbours) or probabilistic (*e.g.* minimal entropy, inter-class t-test) assumptions over the data. Automated thresholding methods can perform reasonably well on clustered and well-separated image intensities but their application to low signal-to-noise (SNR) and contrast-to-noise (CNR) ratio images is not trivial (see fig. 2.1). Thus, more complex image segmentation methods that explicitly model the observed intensities (*e.g.* number of classes, probabilistic distribution of image noise), their spatial localisation and smoothness are a necessary requirement for robust multi-class anatomical segmentation.

2.1.1 Adult brain segmentation

A wide range of image analysis methods have been applied to brain segmentation. One potential broad taxonomy of methods would be to divide them into: methods based on morphological operations (Mangin et al., 1995), edge detection (Tang et al., 2000), fuzzy clustering (Pham, 2002b; Wang and Fei, 2009)

Figure 2.1: (Left) An example image with a low SNR, (Centre) the expected segmentation with each class beeing assigned a different colour, (Right) the resultant segmentation after thresholding.

and probabilistic models (Wells III et al., 1996; Van Leemput et al., 1999b; Zhang et al., 2001; Ashburner and Friston, 2005).

Methods based on morphological operations, like the one proposed by Mangin et al. (1995), normally start by classifying the image intensity using a series of manually or automatically defined thresholds, roughly separating the brain tissue into several classes. Due to the presence of noise and intensity outliers in the image, a series of morphological steps, compromised of an erosion step, followed by a volume based marker selection, segmentation reconstruction (constrained 3D dilation) and ventricle filling using an opening operation, are used. On the other side, edge detection based methods, like the one proposed by Tang et al. (2000), explore the multi-scale nature of human anatomy by performing a sequence of edge detection and region growing steps at iteratively finer scales. These two classes of methods are highly heuristic and very dependent on the image quality, patient anatomy and the presence of MRI intensity inhomogeneity, resulting most of the times in a low segmentation performance. Nonetheless, some conceptual ideas proposed in Mangin et al. (1995) and Tang et al. (2000) are currently being used for sulcal localisation and matching and also in multi-scale feature analysis.

The third class of methods are based on a multi-class fuzzy clustering framework (Pham and Prince, 1999). Fuzzy clustering methods aim at minimising an energy functional compromised of a data clustering term and a smoothly varying gain function to compensate for the existence of MRI intensity inhomogeneities. Even though fuzzy clustering methods do not explicitly model the noise characteristics, the use of a weighted L_2 norm between the observed intensity and the cluster centre makes this class of methods more robust than binary classifiers. The method proposed Pham and Prince (1999) minimises the following energy function:

$$E_{\text{FCM}} = \sum_i \sum_k u_{i,k}(x_i - g_i c_k)^2 + \lambda_1 \sum_i \sum_r (\mathbf{d}_r * g)_i^2 + \lambda_2 \sum_i \sum_r \sum_s (\mathbf{d}_r * \mathbf{d}_s * g)_i^2 \qquad (2.1)$$

where $u_{i,k}$ is the fuzzy membership function, x_i is the data, c_k is the current cluster centroid estimate, g_i is a smoothly variant scaler field used to correct the observed intensity inhomogeneity, $\mathbf{d}_r$ and $\mathbf{d}_s$ are differential operators used to ensure that g_i is spatially smooth and slowly varying. This model was then expanded (Pham, 2002a) to accommodate spatial constraints by adding an MRF-like energy function

over neighbourhood discontinuities to Eq.2.1

$$E_{\text{RFCM}} = E_{\text{FCM}} + \beta \sum_i \sum_k u_{i,k} \sum_{j \in N_i} \sum_{l \in \{1 \ldots k\}} u_{j,l} \tag{2.2}$$

where N_i is the set of voxels neighbouring voxel i. Finally, FANTASM (Pham, 2002b) was introduced by combing all the previous improvements and complementing them with a noise tolerance term. This change, while minor, had a great impact on the performance of the algorithm by not only adding extra robustness to noise and but also by increasing the tolerance to image artefacts and outlier intensities. In order to add time-consistency to the brain segmentation, Xue et al. (2006) introduced the CLASSIC algorithm (`http://www.rad.upenn.edu/sbia/projects/4d_segmentation.html`). CLASSIC is an extension of FANTASM to longitudinal data by using a longitudinal image registration algorithm in conjunction with the introduction of a longitudinal-smoothness term in the energy functional.

According to the above described taxonomy, the last class of methods are the probabilistic mixture models. These models, normally fitted using the expectation maximisation (EM) algorithm, form the basis of several image segmentation methods (Wells III et al., 1996; Van Leemput et al., 1999b; Zhang et al., 2001; Ashburner and Friston, 2005). These EM-based image segmentation algorithms were shown to be among the most accurate and robust (Klauschen et al., 2009) for brain tissue segmentation. More specifically, the maximum likelihood (ML) or maximum a posteriori framework normally takes the form:

$$\hat{\mathbf{\Phi}} = \arg\max_{\hat{\mathbf{\Phi}}} f(\mathbf{y} \mid \mathbf{\Phi}) f(\mathbf{\Phi}) \tag{2.3}$$

where $\mathbf{\Phi}$ are the model parameters, $\mathbf{y}$ are the observed intensities, $f(\mathbf{y} \mid \mathbf{\Phi})$ is the probability of observing the data given a set of current estimates of model parameters and $f(\mathbf{\Phi})$ is the *a priori* distribution of the model parameters. For ML techniques, $f(\log(\mathbf{\Phi}))$ is assumed to be uniformly distributed, where all the $\mathbf{\Phi}$ components are positive. This class of methods were introduced for brain segmentation by Wells III et al. (1996), which segments the brain into three main tissue types (white matter, grey matter and cerebrospinal fluid). Each class is modelled as a normal distribution after log transformation and a Gaussian distributed bias field model is used to correct for INU. Later on, Van Leemput et al. (1999b) added a spatial consistency model based on a Markov Random Field (MRF), explicit modelling of the INU with polynomial basis functions, and some prior information about the brain anatomy to initialise and locally constrain the segmentation. This was achieved by modifying Eq. 2.3 as:

$$\hat{\mathbf{\Phi}} = \arg\max_{\hat{\mathbf{\Phi}}} f(y_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y) f(\mathbf{z}_i = \mathbf{e}_k) \tag{2.4}$$

where $f(\mathbf{z}_i = \mathbf{e}_k)$ encodes the spatial smoothness (MRF) term and the anatomical atlas. This algorithm, later called EMS, is available at `http://mirc.uzleuven.be/MedicalImageComputing/downloads/ems.php`. This model was then extended (Van Leemput et al., 2001) to deal with the presence of outlier observations in order to automatically segment WM lesions in patients with multiple

sclerosis. This updated framework uses an M-estimator with an extra term that is equivalent to adding an uniformly distributed outlier class, thus reducing the influence of outlier samples in the overall likelihood. PV modelling was later integrated into this probabilistic framework (Van Leemput et al., 2003) by using a sub-voxel mixel distribution. Due to the number of degrees of freedom of the model and the ill posed nature of separating noise and PV effects, this model becomes unstable in images with a low SNR. More recently, Ashburner and Friston (2005) combined image registration with tissue classification, and bias field correction in an elegant unified framework, part of the SPM (Statistical Parametric Mapping) package (`http://www.fil.ion.ucl.ac.uk/spm/`). This framework models PV by using more than one Gaussian per class and iteratively registers the population anatomical priors while segmenting the image in a fully generative framework. Even though it does not have a spatial consistency term, the combination of an iteratively aligned population atlas and the PV modelling results in highly accurate segmentations.

Despite these advances, the accuracy of segmenting fine structures is still limited by problems with INU, PV, noise, image artefacts, limited resolution and the great degree of natural variability. These problems can lead to an incorrect delineation of problematic areas like PV-corrupted grey matter folds, resulting in non-optimal segmentations. The use of prior knowledge may also cause problems in areas that have an high degree of natural variability, as the prior information is representative of a sample of a normal population and might not describe a particular subject. The use of probabilistic priors becomes more problematic when an atlas derived from a normal population is used to segment patients with different anatomical or pathological characteristics. Furthermore, current methods also ignore the information about cortical shape. Thus, a new methodological approach that combines anatomical knowledge about the shape of the cortex with a more physically plausible and PV aware probabilistic imaging model is warranted.

2.1.2 Neonatal segmentation

Babies born very preterm (less than 32 weeks of gestation) are at increased risk of a range of cognitive and learning problems that become more frequent with lower gestations at birth (Marlow et al., 2005; Mathur and Inder, 2009; Ment et al., 2009). Although survival rates of preterm infants have improved significantly over the last few decades, this has not been accompanied by a reduction in rates of neurodisability, or improvement in the cognitive outcome for survivors. As a result, a large proportion of very preterm children have disabilities and special educational needs with consequent high societal costs (Mangham et al., 2009). Following very preterm birth, brain development occurs outside the normal protective environment, exposing the infant to a range of external stimuli and problems of homeostasis. Changes seen on MRI at term equivalent age are likely to represent the amalgamation of cellular injury and disturbance of normal brain development (Khwaja and Volpe, 2008; Volpe, 2009). Studies using advanced magnetic resonance techniques over the past decade have begun to identify differences in size and structure of the brains of preterm infants imaged at term compared with term born controls and correlation with neurodevelopmental outcome at 2 years (Boardman et al., 2010). As stated before, fundamental to performing volumetric and morphometric studies is the ability to classify different brain

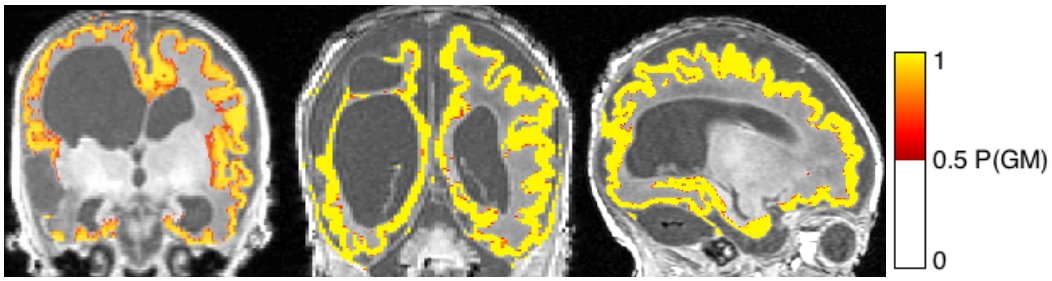


Figure 2.2: Three highly pathological neonatal subjects overlayed with a probabilistic automated GM segmentation.

tissues. In contrast to adults, the segmentation of the neonatal brain is complicated due to a combination of: low-signal-to noise ratio; increased PV as a result of adapting the resolution to the smaller neonatal head; and the existence of both natural and pathological hypo- and hyper- intensities. In addition, there is substantial natural and pathological variability due to the effects of prematurity and the spectrum of preterm brain injury as seen in Fig. 2.2 (Kapellou et al., 2006; Rutherford et al., 2010).

Maturational dependent intensity differences are well-known between the adult and neonatal brain. There is a dynamic natural developmental variability which arises due in part to the receding germinal matrix and progressive myelination, which manifests as an apparent reversal of the signal intensities of grey and white matter (GM/WM) on neonatal T1- and T2-weighted MRI. As myelination proceeds during the first months of life, contrast between the two tissue types progressively changes until an adult intensity pattern emerges around two years of age. The complexity of the cortical surface also increases rapidly over the period between 20-40 weeks gestational age, corresponding to the preterm period raising the possibility that cortical surface analysis may provide an independent predictive biomarker of neurological outcome (Dubois et al., 2008).

As a result of the paramount role of segmentation, a number of authors have produced techniques specifically for neonatal MRI, primarily by adapting and enhancing well-established techniques in the adult brain. Prastawa et al. (2005) developed a technique based on the previously described canonical expectation maximisation (EM) method of Van Leemput et al. (1999b) by explicitly modelling additional classes of white matter to account for the myelination process. The addition of this extra tissue class for myelinated white matter, which they found primarily in the deep brain, made this technique a useful possibility for studies of myelination changes during infancy. The authors subsequently applied their technique to larger volumes of data in Gilmore et al. (2007). Xue et al. (2007) proposed a series of improvements to neonatal segmentation specifically for the cortex, combining the brain extraction technique with deep grey matter structure removal and particularly highlighting the problem of misclassified PV pixels (*e.g.* unlikely white matter pixels appearing between dark cortical grey-matter and light cerebral spinal fluid on T2-weighted MRI) through morphological operations without implicitly modelling PV within the EM procedure. This addition is specific to neonatal MR contrast, but equivalent to the corresponding adult problem of unlikely grey matter pixels appearing between white matter and ventricular CSF. The authors subsequently used the segmentation results to extract and analyse the cortical

surface using a level-set routine for further processing. In the absence of a neonatal atlas, the authors initialised the segmentation with the results of a k-means clustering. Both Prastawa et al. (2005) and Xue et al. (2007) applied their techniques to infants with normal anatomical appearance. More recently, Shiee et al. (2011) addressed the problem of the segmentation of cases that are far from the space of the priors, specifically adult ventriculomegaly, by allowing an iterative 'relaxation' of the anatomical priors by deriving them from a Dirichlet distribution, an approach likely to be useful for the diverse anatomy of neonatal cohorts. This procedure thus allows pixels far from the prior atlas to become progressively incorporated in the classification allowing the segmentation of pathological cases. An alternative segmentation methodology was proposed by Weisenfeld and Warfield (2009), employing template propagation and fusion to estimate the most likely tissue classifications (including classes for unmyelinated white matter and subcortical grey matter). In this algorithm, sub-structures that are not identified in a three or four class EM routine may be extracted and analysed and the method mitigates the effect of low signal and contrast to noise ratios. However, template driven segmentation methods rely on the availability and registration of well-defined templates in order to propagate knowledge of a particular population to a patient specific space. This is problematic in pathological cases, as their anatomy may be markedly different from the one derived from a normal population atlas and the process may not be mitigated by registration.

More recent advances in segmentation methodology have seen a number of atlas-driven segmentation methods proposed as a result of a number of groups developing and making available specific neonatal atlases, averaged over a number of subjects: Knickmeyer et al. (2008) developed a three-class atlas-based approach from birth to two years, analysing population changes over this period and thus not distinguishing between myelinated and unmyelinated white matter; Oishi et al. (2011) made use of diffusion-weighted data to produce a multi-contrast neonatal brain atlas; while Kuklisova-Murgasova et al. (2011) have made available a comprehensive neonatal atlas from 29-44 weeks generated from T2-weighted data with intensity classes for grey matter, white matter and CSF and spatial sub-classes for sub-cortical grey matter, cerebellum and brainstem. Song et al. (2007) proposed a simpler intensity-based classification method based on k-means and augmented by a population atlas and Yu et al. (2010) proposed a Parzan windows based Hidden Markov Random Field algorithm, optimised by an Expectation-Maximisation algorithm, again augmented by a population atlas.

Notwithstanding their performance in morphologically normal subjects, all these methods have problems in severely diseased patients as the propagation of the atlas/templates by image registration is not able to cope with the variability. Furthermore, the image contrast requirements of the above described algorithms restrain their applicability to T2-weighted images which have a much lower out-of-plane resolution. It is thus important to research methods that enable the segmentation of pathological datasets in a robust and accurate manner.

2.2 Brain extraction and structural parcellation

Extracting the brain and dividing it into multiple regions-of-interest (Hammers et al., 2003) reduces the complexity of image segmentation tasks and enables the analysis of specific anatomical structures. As

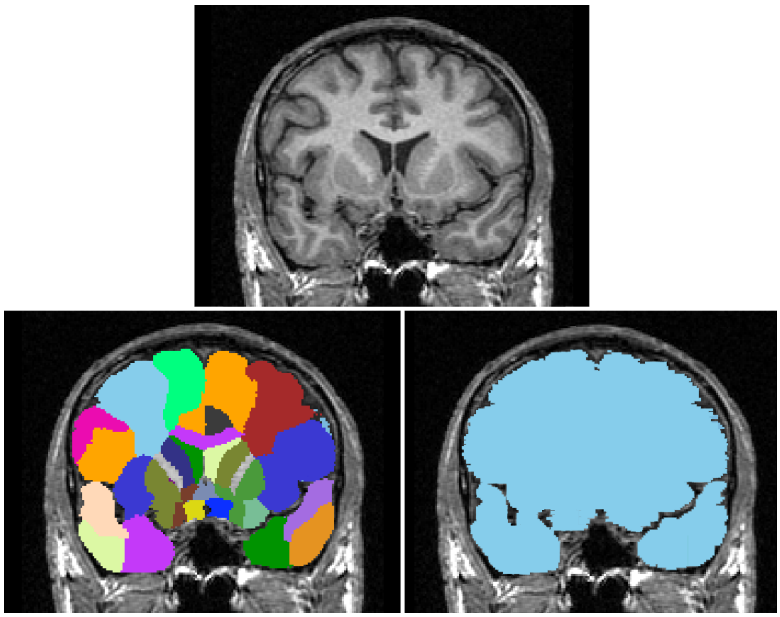


Figure 2.3: Example of a structural parcellation (bottom left) from the Hammer's Atlas (Hammers et al., 2003) and its corresponding brain extraction (bottom right) overlayed on a T1-weighted image.

structures that are anatomically distinct do not necessarily differ in their signal properties and can be composed of more than one tissue type, this process is usually difficult to automate. Three main classes of algorithms have been introduced: surface and shape based models (Smith, 2002; Segonne et al., 2004; Patenaude et al., 2011), probabilistic segmentations with population priors (Fischl et al., 2002), and multi-atlas segmentation propagation and fusion (Aljabar et al., 2009; Artaechevarria et al., 2009; Heckemann et al., 2010).

Surface-/contour-based methods try to fit a parameterised surface to the region of interest by adapting it to the patient specific morphology while maintaining certain shape characteristics. For example, the brain extraction tool (BET) algorithm (Smith, 2002) starts by thresholding an image of the head into background and foreground followed by a centre-of-gravity (COG) localisation. After this COG is found, a spherical mesh with a radius smaller than the brain is initialised and evolved outwards until in reaches a low intensity area (CSF) while maintaining a certain degree of smoothness. Even though the smoothness term supposedly ameliorates the problem of having the surface leaking to the outside of the brain, the robustness of BET for large scale studies is limited (Rex et al., 2004). Another example of this kind of algorithms is the hybrid watershed algorithm (HWA) from the FreeSurfer package (Segonne et al., 2004). This algorithm initialises the mesh by running a watershed algorithm with a split-and-merge strategy that fuses several candidate watershead basins into a initial brain segmentation. A mesh is generated from this initial brain segmentation and evolved toward the brain edge according to the local intensity patterns. The evolution of this mesh is constrained by *a priori* knowledge about the shape of the brain, resulting in a robust brain extraction algorithm. More recently, Patenaude et al. (2011) presented a method for parcelating the brain into its multiple composing structures. This method used an active shape and appearance model formulated within a Bayesian framework that allows the

probabilistic relationships between shape and intensity to be fully exploited. A similar method, but for hippocampal segmentation, is intuitively depicted in Fig. 2.4.

The second class of methods used for structural parcellation is similar in nature to the probabilistic models that were previously described for tissue segmentation. Methods like the one presented by Fischl et al. (2002) register a probabilistic population atlas with 37 tissue classes and an associated spatially variant MRF that characterises not only the classes that should be neighbouring each other at each spatial location in the image but also the directionality of this neighbouring. Due to the presence of the MRF, the algorithm becomes robust to noise and image artefacts. A similar method (Lötjönen et al., 2011) was developed for single-structure parcellation, where a series of hippocampal atlases are propagated through composition of deformation fields, followed by a PV aware EM classification (see Fig 2.5). As both this strategies are based on population atlas, the algorithm is not able to cope with the presence of outlier morphologies, resulting in low Dice scores for certain structures when compared to other methodologies.

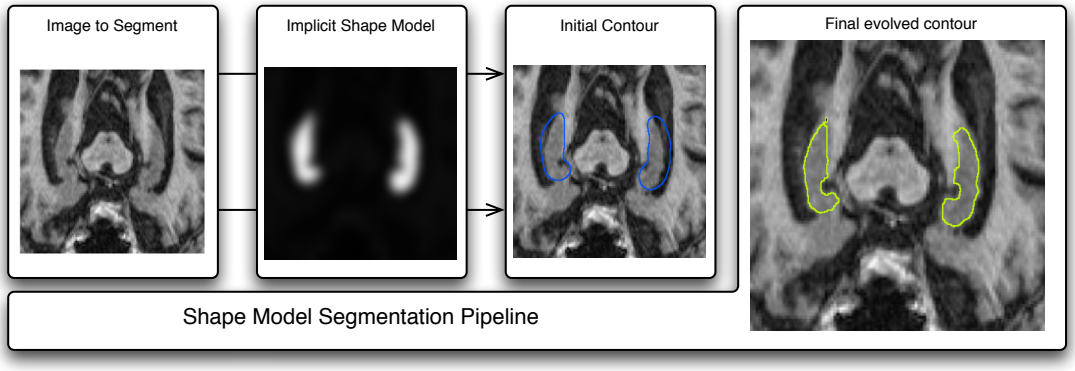


Figure 2.4: Depiction of a contour based hippocampal segmentation. A implicit shape model is first aligned with the target structure. An intial contour is then evolved in order to match the edges of the structure of interest, whilst respecting the model's curvature and shape contrains.

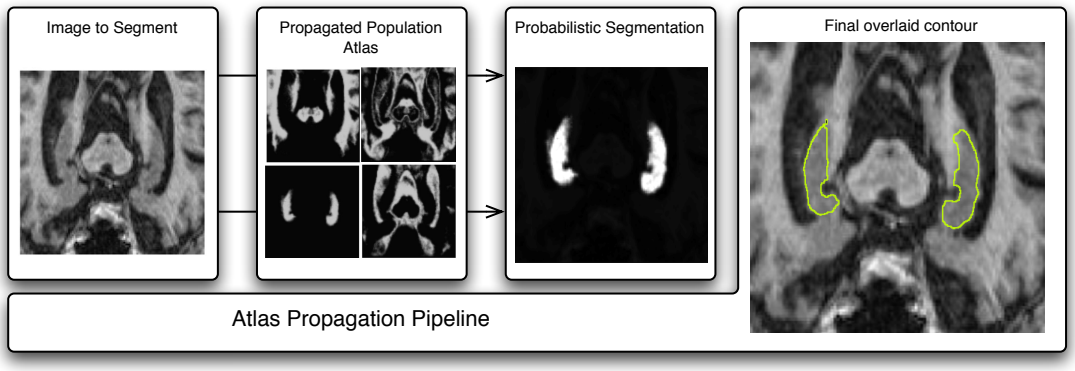


Figure 2.5: Sumary diagram of a atlas propagation based structural parcellation. A population atlas and its composing tissue priors are propagated to an unseen subject. These propagated priors, that have a class specific for the target structure of interest, are used as the input to an EM probabilsitic segmentation framework in order to obtain the final segmentation.

The last class of methods that is becoming increasingly popular due to their high accuracy is multi-atlas segmentation propagation and fusion. Due to advances in both speed and accuracy of registration techniques, segmentation propagation is becoming an increasingly popular segmentation and parcellation technique in medical imaging research. The method is summarised in Fig. 2.6. Recent studies have shown that the accuracy and robustness of the segmentation can be greatly improved by combining multiple segmentations from a library/database of atlases (Rohlfing et al., 2004b). Each image from the database, when registered to the image of interest, can be considered as an independent classifier. Conventionally, each classifier is combined assuming equal weights, a methods known as voting fusion (Xu et al., 1992). As some classifiers will be more accurate than others, or in the case of segmentation propagation some registrations will be more accurate than others, more sophisticated techniques to quantify classifier performance have been developed. The classifiers can be combined according to their performance on a training set (Lam and Suen, 1995), or by estimating its performance on a feature or metric space (Woods et al., 1997).

On a segmentation propagation perspective, Aljabar et al. (2009) noted that the presence of outlier classifiers (due to bad registration) was causing a bias in the estimated segmentation and consequently a degradation of the final results in a voting scheme. They proposed to use the global normalised cross-correlation metric as a performance estimator in order to select the optimal classifiers for the voting scheme. While for simple shapes a global metric might be sufficient, for objects with complex morphometry, the size of the database has to increase dramatically to be able to characterise the population's morphometric variability. Artaechevarria et al. (2009) proposed a set of local and global performance estimators based on image similarity metrics like the global normalised cross correlation (GNCC), global mean square difference (GMSD), global mutual information (GMI) and the corresponding local versions of the metrics, LNCC, LMSD and LMI respectively. More recently, Yushkevich et al. (2010) proposed a modified version of the LNCC metric using a ranking scheme and Collins and Pruessner (2010) used a GMI metric combined with a registration to a group-wise space in order to reduce computational cost. All in all, voting methods assume that the original manual segmentation is correct and that classifier performance will only depend on image similarity, which is not true. Instead of using an image similarity derived performance metric, an algorithm named STAPLE (Simultaneous Truth and Performance Level Estimation) was proposed by Warfield et al. (2004) as a novel way to estimate the performance parameters of a classifier and consequently obtain the most probable classification. This algorithm estimates the performance parameters by comparing each classifier to a consensus classifier, in an iterative manner. Leung et al. (2010) then introduced the same ranking concept as in (Aljabar et al., 2009) in a STAPLE framework, resulting in an improved segmentation accuracy. However, as only global metrics were introduced, these methods still suffer from the same above described problems (*e.g.* complex morphometry, local matching). Also, these global metrics are dependent from the ROI where they are calculated, *i.e.* they can be calculated on the full image or on the ROI close to the object of interest, and are not robust to INU uniformity (INU) in MRI images.

It is thus important to research how to integrate a local similarity metric into a STAPLE formulation,

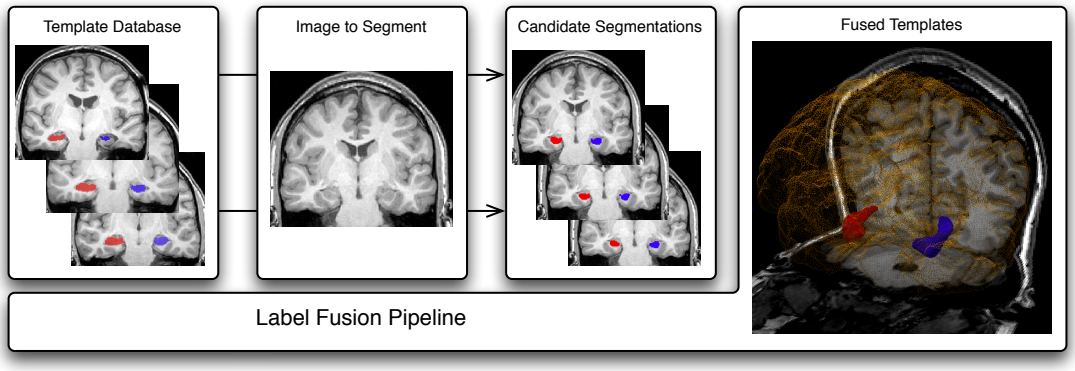


Figure 2.6: Sumary diagram of a multi-atlas segmentation propagation and fusion framework. A set of atlases are first registered to the target image. The segmentations associated to each atlas are then propagated to the target image using the same transformation and subsequentely fused into a final segmentation.

as it would combine the advantages of estimating segmentation performance and local image similarity in an unified and statistically rigorous framework.

2.3 Cortical Thickness

The automatic extraction of thickness measurements from anatomical structures has the potential to provide a biomarker for diagnosis and disease progression. However, the reliable extraction of sub-voxel accurate measurements of thickness from probabilistic segmentations is still an unsolved problem. Thickness estimation methods can be separated into surface-based, registration based and voxel-based techniques.

Surface based methods (Fischl and Dale, 2000; Kim et al., 2005), like FreeSurfer, follow a series of steps to normalise and segment the image, followed by a surface fitting step. In FreeSurfer, for example, processing involves intensity normalisation, registration to Talairach space, skull stripping, segmentation of white matter, tesselation of the WM boundary, smoothing of the tesselated surface and automatic topology correction. The tesselated surface is then used as the starting point for a deformable surface algorithm to find the WM and then the pial boundary. For each point on the tesselated WM surface, the cortical thickness is calculated as the average of the distance from the WM surface to the closest perpendicular point on the pial surface and from that point back to the closest perpendicular point on the WM surface (see Fig 2.7).

On the other side, the CRUISE method (Kim et al., 2005), another surface based cortical thickness estimation algorithm, has an implicit coupling of these two surfaces, thus resulting in a more geometrically stable estimate of thickness (see Fig 2.8).

However, surface based cortical thickness methods have several drawbacks. The fitting of a triangulated mesh to the region of interest makes this class of methods computationally expensive, especially if a requirement is to maintain topological constraints. Surface based methods produce very robust and noise free results due to the implicit curvature and smoothness constraints and the use of regional information, but the parametrisation of the surface can be complex and curvature constraints can bias the thickness

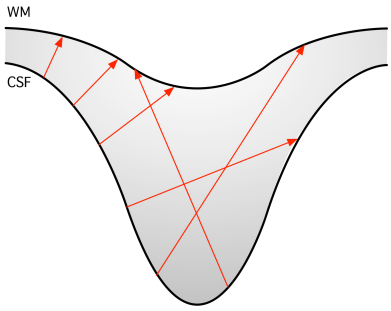


Figure 2.7: Diagram showing the perpendicular nearest point search from the WM/GM boundary to the pial surface in Freesurfer. Note the directionality problems in highly convoluted structures.

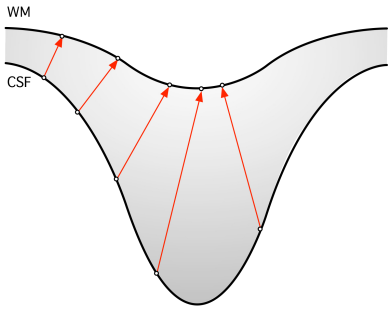


Figure 2.8: The CRUISE method, when compared to Freesurfer, reduces the geometrically derived estimation problems by coupling the two cortical surfaces together.

measurements (Scott et al., 2009). The thickness measurement can also be biased by the choice of the smoothness parameters as these will impact the ability of the algorithm to fit a mesh to highly convoluted structures (Scott et al., 2009).

Registration based cortical thickness algorithms use the notion of coordinate mapping to calculate cortical thickness. One example of this approach is the work byDas et al. (2009), where a WM segmentation of the brain is deformed and mapped to a probabilistic segmentation containing both WM and GM segmentations. Thus, the distance necessary to map the WM segmentation to the (WM+GM) segmentation corresponds to the thickness of the cortex. In this method, if the initial WM segmentation is topologically correct, then the final thickness estimation would also be topologically correct. This approach has problems regarding the bias/variance tradeoff between the registration regularisation term and the image similarity term. An higher regularisation ensures smoothness in the thickness measurements but impedes the expansion of highly collapsed gyri, while a higher image similarity results in noisy measurements of thickness and noisy sulci localisation.

Voxel-based methods on the other side extract the value of thickness directly from the voxel grid and are computationally very efficient, however, their accuracy is critically dependant on the image resolution and the quality of the segmentation. Furthermore, topological problems might occur in highly convoluted areas because of the limited knowledge about spatial information. Overall, voxel-based methods can be clustered into 3 subgroups: mathematical morphology (Lohmann et al., 2003), partial differential equation (Hutton et al., 2008; Jones et al., 2000; Yezzi Jr and Prince, 2003; Rocha et al., 2005; Acosta

et al., 2009) and line integral based (Aganj et al., 2008).

Mathematical morphology based methods use a combination of skeletonisation and region growing techniques in order to calculate the minimal Euclidean distance between points. These methods, intuitively represented in Fig. 2.9 by fitting spheres into the cortical sheet, work on a binarised version of the segmentation thus having their accuracy limited by the voxel size.

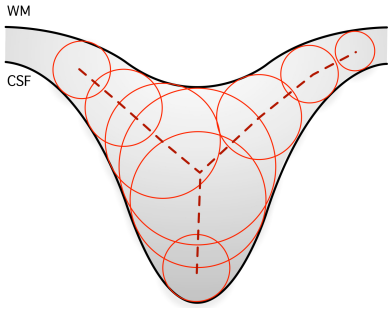


Figure 2.9: Diagram showing the mathmatical morphology based CTE. Note the shpere fitting problems in highly curved areas.

Partial differential equation (PDE) based methods solve the Laplace equation between the inner and outer surfaces as if they were charged conductors, resulting in isopotential electric field lines between them. The thickness is then equal to the sum of the lengths of the normals to these isolines (see Fig. 2.10). The normals to the isolines that go from the inner to the outer surfaces area are guaranteed to be diffeomorphic, as they are bijective and never cross. However, this method requires both cortical surfaces to be homotopic to a sphere in order to avoid saddle points in the Laplacian field and collisions on the advancing front. Even though this is a necessary condition, the Laplace equation based methods are used without enforcing these topological constraints (Acosta et al., 2009).

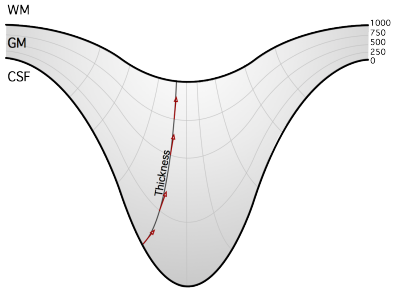


Figure 2.10: Diagram showing the Laplace equation based CTE. Note that this method does not have problems in highly curved areas as show in Fig. 2.9 . Also, it guarantees smoothness and bijectivity in the CTE metric.

Finally, line integral based methods calculate thickness of the structure of interest by finding the direction that minimises the line integrals over its probabilistic segmentation at each position of the 3D volume (see Fig. 2.11). The accuracy of this method is very high for simple shapes because it works on the probabilistic segmentation and not on a binarised grid, however, the choice of stopping criteria becomes critical in highly convoluted areas like the cerebral cortex.

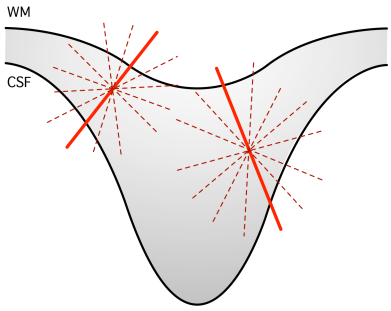


Figure 2.11: Diagram showing the estimation of cortical thickness though minimal line integration at each point in the cortex.

A comparison of voxel and surface based cortical thickness estimation methods, presented by Clarkson and Cardoso et al. (2011), showed that the accuracy of surface and voxel based methods is similar for most applications, with lower test-retest variability for surface based methods but higher sensitivity for voxel based methods (Clarkson et al., 2011). While most cortical thickness to date studies have been done cross-sectionally (observation of a representative subset of a population at one specific point in time), recent studies (Desikan et al., 2008; Engvig et al., 2010; Sabuncu et al., 2011) have demonstrated the importance of exploring disease progression and their imaging signatures through the use of longitudinal studies (repeated observations of the same variables over long periods of time). The extraction of 4D consistent measurements of thickness from anatomical structures is of interest in various diseases such as Alzheimer's and Huntington's disease, having the potential to provide a biomarker for diagnosis and neurodegeneration (Holland et al., 2009). Nonetheless, the reliable extraction of 4D consistent and sub-voxel accurate measurements of thickness from probabilistic segmentations is still an unsolved problem.

In order to provide accurate longitudinal measurements, one requires that the measurement of thickness is performed in a consistent manner on all time points, even in the presence of atrophy, sulcal or gyral collapse and overall morphological changes (see Fig. 2.12). However, from the 3 main steps of any CTE algorithm:

1. Segmenting the image
2. Finding an implicit or explicit correspondence between both sides of the cortex
3. Measuring the thickness according to some metric

altering steps 1 or 3 is a non ideal approach as it will directly constrain what is being measured. This can reduce the statistical significance of difference measurements between groups, leading to increased sample sizes. Instead, in order to maintain the sensitivity of the measurement intact but still achieve temporal consistency, one can constrain only the direction of the measurement by changing step 2. This solution will be explored later in this thesis.

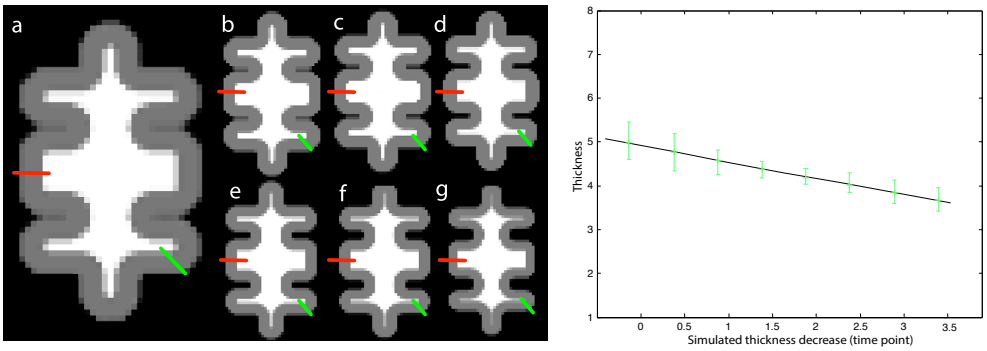


Figure 2.12: (Left) Digital phantom with a simulated atrophic process. (Right) The expected average cortical thickness in black and estimated thickness in green. Under ideal conditions, the atrophic process should not invert the direction of cortical thickness changes.

Chapter 3

Tissue Segmentation

3.1 LoAd: a locally adaptive cortical segmentation algorithm

As stated in the introduction chapter, segmentation of MRI images in elderly and demented patients is normally hindered by many factors like INU, PV, noise, image artefacts, limited resolution and the great degree of natural variability. Thus, new algorithms have to be developed to address these problems. Even though there are a multitude of segmentation algorithms already available, I'll demonstrate that small modifications in the segmentation pipeline can have a significant impact on both the bias introduced by the population's anatomical atlas, the characterisation of PV and tissue fraction and even on the quality of the cortical delineation in the presence of noise. I'll also demonstrate that these modifications will impact, not only on the accuracy of the segmentation, but also on the estimation of cortical thickness.

3.1.1 Derived publications and methodological sumary

- Andrews, K., Modat, M., Macdonald, K., Yeatman, T., **Cardoso, M. J.**, Leung, K., Barnes, J., Fox, N., Ourselin, S., Schott, J., Cross-sectional cerebral volumetric differences and associations with estimated time to age-at-onset in familial Alzheimer's disease: Findings from the DIAN study (2012) Alzheimer's & Dementia 8 (4), p22-p23
- **Cardoso, M. J**., Clarkson, M. J., Ridgway, G. R., Modat, M., Fox, N. C., Ourselin, S., The Alzheimer's Disease Neuroimaging Initiative, LoAd: A locally adaptive cortical segmentation algorithm, (2011), Neuroimage
- Clarkson, M. J., **Cardoso, M. J.**, Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., Fox, N. C., Ourselin, S.,: A Comparative Study of Voxel and Surface Based Cortical Thickness Methods in Frontotemporal Dementia , (2011), AAIC
- **Cardoso, M. J.**, Clarkson, M. J., Modat, M., Ridgway, G. R., Ourselin, S. Locally weighted Markov random fields for cortical segmentation, (2010), IEEE ISBI
- **Cardoso, M. J.**, Clarkson, M. J., Modat, M., Ridgway, G. R., Fox, N.C., Ourselin, S: Improved Maximum a Posteriori Cortical Segmentation by Iterative Relaxation of Priors. (2009) MICCAI

The probabilistic segmentation framework from Van Leemput et al. (1999b) is improved with three novel modifications: reduction of the influence of the priors in an anatomically coherent way, improve-

Figure 3.1: Segmentation of a BrainWeb T1-weighted dataset with 3% noise and 20% INU: (Left) BrainWeb ground truth segmentation; (Centre) MAP with MRF but without the proposed improvements; (Right) Proposed method.

ment of the PV estimation and improvement of the delineation of deep sulci and gyri (Fig.3.1). Both the solution of the EM algorithm and the information derived from a geodesic distance function are used to locally modify the priors and the weighting of the MRF, enabling the detection of small variations in intensity while maintaining robustness to noise. An MRF energy matrix derived from the anatomical properties of the brain is used to add topological and shape knowledge to the MRF. Although full topological correctness is not ensured, the proposed MRF energy matrix improves the topological characteristics of the segmentation and reduces the PV layer thickness, making it more in line with the theoretical anatomical limit. The implicit modelling of PV and the reduction of the PV layer thickness obviates the need for an empirical threshold to distinguish between pure and mixed voxels and eases the problem of achieving subvoxel accuracy when calculating the cortical thickness.

3.1.2 Methods

Intensity Model and MRF regularisation

Starting from the image model developed by Van Leemput et al. (1999b), let $i \in \{1, 2, \cdots, n\}$ index the n voxels of an image domain. For coregistered multimodal datasets, the logarithm of the intensities are used as feature vectors $y_i \in \mathbb{R}^m$; here, for simplicity, I assume unimodal data with $m = 1$. Let $\mathbf{z}_i$ denote the tissue type to which voxel i belongs. For K tissue types, $\mathbf{z}_i = \mathbf{e}_k$ for some k, $1 \leq k \leq K$ where $\mathbf{e}_k$ is a unit vector with the kth component equal to one and all the other components equal to zero.

The log transformation of the image intensities makes the multiplicative bias field additive, simplifying the optimisation procedure by enabling the existence of a linear least square solution for the coefficient optimisation and ameliorating problems with numerical stability (Van Leemput et al., 1999b). An example of the original and log transformed intensity histogram can be seen in figure 3.2. Thus, as in Van Leemput et al. (1999a), the INU bias field is represented by a linear combination $\sum_{j=1}^{J} c_j \phi_j$ of J smoothly varying basis functions $\phi_j(x)$, where x denotes the spatial position and $\mathbf{c} = \{c_1, c_2, ..., c_j\}$ denote the bias field parameters. Note that the complexity and accuracy of the fit of the bias field might not hold for high field MRI images.

For mathematical convenience and similarly to (Wells III et al., 1996), (Van Leemput et al., 1999b) and (Zhang et al., 2001), the intensity of the voxels that belong to class k is assumed to be normally distributed after log transformation with mean μ_k and standard deviation σ_k grouped in $\boldsymbol{\theta}_k = \{\mu_k, \sigma_k\}$. Let $\boldsymbol{\Phi}_y = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_K, \mathbf{c}\}$ represent the overall model parameters.

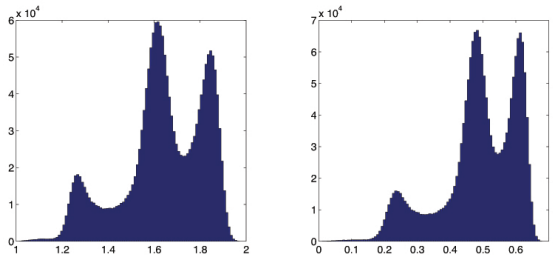


Figure 3.2: Two intensity histograms from the brain region of a T1 MRI image: (left) the original intensities normalised between 1 and 2 and (right) after log transformation of the normalised intensities.

Defining $\mathbf{\Phi}_y$ as the model parameters, the overall probability density for y_i is

$$f(y_i \mid \mathbf{\Phi}_y) = \sum_k f(y_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y) f(\mathbf{z}_i = \mathbf{e}_k) \tag{3.1}$$

with

$$f(y_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y) = G_{\sigma_k}\Big(y_i - \mu_k - \sum_j c_j \phi_j(x_i)\Big) \tag{3.2}$$

where $G_{\sigma_k}(\,)$ denotes a zero-mean normal distribution with standard deviation σ_k. Equation 3.1 can be seen as a mixture of normal distributions.

Thus, by assuming statistical independence among voxels, the overall probability density for the full image can be given by

$$f(\mathbf{y} \mid \mathbf{\Phi}_y) = \prod_i f(y_i \mid \mathbf{\Phi}_y) \tag{3.3}$$

The Maximum Likelihood (ML) parameters for $\mathbf{\Phi}_y$ can be found by maximisation of $f(\mathbf{y} \mid \mathbf{\Phi}_y)$, giving the following update equations for the model parameters:

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^n p_{ik}^{(m+1)} \Big(y_i - \sum_{j=1}^J c_j^{(m)} \phi_j(x_i)\Big)}{\sum_{i=1}^n p_{ik}^{(m+1)}} \tag{3.4}$$

$$\sigma_k^{(m+1)} = \sqrt{\frac{\sum_{i=1}^n p_{ik}^{(m+1)} \Big(y_i - \mu_k^{(m+1)} - \sum_{j=1}^J c_j^{(m)} \phi_j(x_i)\Big)^2}{\sum_{i=1}^n p_{ik}^{(m+1)}}} \tag{3.5}$$

where

$$p_{ik}^{(m+1)} = \frac{f\Big(y_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y^{(m)}\Big) f\Big(\mathbf{z}_i = \mathbf{e}_k\Big)}{\sum_{j=1}^K f\Big(y_i \mid \mathbf{z}_i = \mathbf{e}_j, \mathbf{\Phi}_y^{(m)}\Big) f\Big(\mathbf{z}_i = \mathbf{e}_j\Big)} \tag{3.6}$$

is a weight at the index i and class k and m denotes the iteration number. The estimation of $c_j^{(m+1)}$ is provided by Van Leemput et al. (1999b).

Anatomical priors that incorporate probabilistic information derived from a digital brain atlas are added to the model in order to condition the posterior probabilities and indirectly condition the model parameters. These atlases are brought into correspondence using an affine registration (Ourselin et al.,

2000, 2001) followed by a free-form non-rigid registration algorithm (Modat et al., 2010) [1] and are introduced as a weight π_{ik}, integrated in eq.3.1 by making $f(\mathbf{z}_i = \mathbf{e}_k) = \pi_{ik}$. Equations 3.4, 3.5 and 3.6 remain valid and the initial values for p_{ik}^0, μ_k^0 and σ_k^0 are given by their equations with $c_j = 0$ and $f(y_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y) = 1$.

I assume that all images are previously skull stripped and I initially model the problem with $K = 6$ classes, each one with a corresponding digital atlas prior probability for white matter (WM), cortical grey matter (cGM), deep grey matter (dGM), external cerebrospinal fluid (eCSF), internal cerebrospinal fluid (iCSF) and dura (DU) respectively at every voxel position. These priors are derived from the ICBM Tissue Probabilistic Atlas [2] and are created by merging several priors from several areas together. The procedure used for skull stripping will be described later in this thesis.

The cortical and deep GM are modelled as separate classes to enable thickness calculation over the cortical structures and to enable the segmentation of a broader range of pulse sequences (*e.g.* new quantitative MR sequences that look at iron concentration - R2 and R2* maps (Gelman et al., 1999)), that have differing intensities for dGM and cGM. The distinction between deep and cortical GM and internal and external CSF also enables different topological and connectivity properties to be assigned to each class. For example the iCSF, *i.e.* the CSF within the ventricles, can be next to WM or dGM voxels while the eCSF can only be next to cGM voxels. Finally, the dura class is used to compensate for problematic skull stripping situations.

Unfortunately, the intensity model alone only works in relatively ideal conditions because it classifies the voxels of the image based solely on intensity and on the assumption that neighbouring voxels are independent. This makes the segmentation very prone to noise and image artefacts. Therefore, the model has to be made more robust to noise by augmenting the spatial tissue priors with additional prior knowledge about topology and spatial smoothness. This can be achieved by the using an MRF which assumes that the probability that voxel i belongs to tissue k depends on its first-order 3D neighbours $\mathcal{N}_i$. Using the mean field approximation as described in (Zhang, 1992) and (Van Leemput et al., 1999b), eq. 3.6 becomes

$$p_{ik}^{(m+1)} = \frac{f(y_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y^{(m+1)}) f(\mathbf{z}_i = \mathbf{e}_k \mid p_{\mathcal{N}_i}^{(m)} \mathbf{\Phi}_z^{(m)})}{\sum_{j=1}^{K} f(y_i \mid \mathbf{z}_i = \mathbf{e}_j, \mathbf{\Phi}_y^{(m+1)}) f(\mathbf{z}_i = \mathbf{e}_j \mid p_{\mathcal{N}_i}^{(m)} \mathbf{\Phi}_z^{(m)})} \tag{3.7}$$

with,

$$f(\mathbf{z}_i = \mathbf{e}_k \mid p_{\mathcal{N}_i}^{(m)} \mathbf{\Phi}_z^{(m)}) = \frac{\pi_{ik}\, e^{-\beta_i U_{\text{MRF}}(\mathbf{e}_k \mid p_{\mathcal{N}_i}^{(m)}, \mathbf{\Phi}_z^{(m)})}}{\sum_{j=1}^{K} \pi_{ij}\, e^{-\beta_i U_{\text{MRF}}(\mathbf{e}_j \mid p_{\mathcal{N}_i}^{(m)}, \mathbf{\Phi}_z^{(m)})}} \tag{3.8}$$

where $U_{\text{MRF}}(\mathbf{z}_i \mid p_{\mathcal{N}_i}, \mathbf{\Phi}_z)$ is an energy function dependent on the parameters $\mathbf{\Phi}_z$ and, at this stage $\beta_i = 1 \; \forall \; i$. Due to the possibility of anisotropic voxel size and slice spacing, the interaction between neighbours in each direction should be different. To take this property into account, a connection strength factor s is introduced as $s = \{s_x, s_y, s_z\} = \{\frac{1}{d_x}, \frac{1}{d_y}, \frac{1}{d_z}\}$, where d is the real-world distance between the centre of neighbouring voxels in each direction. This approach leads to higher weights in the MRF

[1] Available from `http://sourceforge.net/projects/niftyreg/`
[2] Available from `http://www.loni.ucla.edu/ICBM/ICBM_Probabilistic.html`

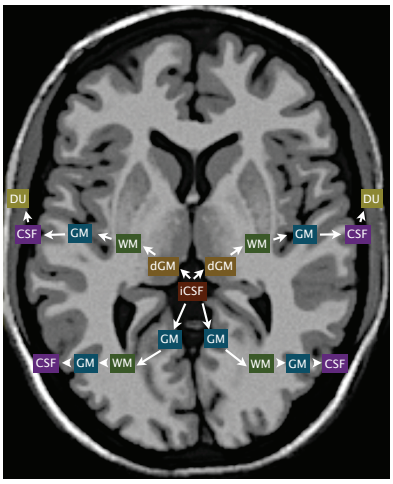


Figure 3.3: MRF class connectivity network.

when voxels are closer together. Under this framework,

$$U_{\text{MRF}}(\mathbf{e}_k \mid p_{\mathcal{N}_i}, \mathbf{\Phi}_z) = \sum_{j=1}^{K} G_{kj} \left(\sum_{l \in \mathcal{N}_i^x} s_x \, p_{lj} + \sum_{l \in \mathcal{N}_i^y} s_y \, p_{lj} + \sum_{l \in \mathcal{N}_i^z} s_z \, p_{lj} \right) \tag{3.9}$$

where $\mathbf{\Phi}_z = \{\mathbf{G}, s\}$, with $\mathbf{G}$ as a K x K matrix with element $\mathbf{G}_{kj}$ containing the transition energy between tissue k and j, and with the MRF neighbourhood system defined as $\mathcal{N}_i = \left\{\mathcal{N}_i^x, \mathcal{N}_i^y, \mathcal{N}_i^x\right\}$.

Although G can be estimated and updated using a mean field theory based approximation, these estimates are only representative of the global image statistics and not of the known brain anatomy. Furthermore, the presence of noise can hamper the correct estimation of the MRF energy matrix. Instead of estimating and updating $\mathbf{G}$ at each iteration, constant values are assumed based on anatomical proprieties of the brain. The MRF class connectivity network is represented in Fig.3.3. The classes connected with arrows are considered neighbouring classes, and the ones that are not connected are considered distant classes. Even though this connectivity matrix is representative of most anatomical neighbouring features, in areas like the ventricle edges, a layer of GM will be assigned to the glial tissue and the PV corrupted voxels in the interface between WM and CSF. This will also happen in areas like the pons. These small anatomical incoherences are related to the constant MRF energy matrix $\mathbf{G}$. A spatially varying MRF energy matrix could be used to spatially change the neighbouring rules, however, this would greatly increase the computational complexity. One should also bear in mind that the neighbouring rules are not a hard constraint. Matrix $\mathbf{G}$ is defined as:

$$\mathbf{G}_{kj} = \begin{cases} 0 & \text{if class } k \text{ is the same as } j \\ \alpha & \text{if class } k \text{ is neighbouring } j \\ \gamma & \text{if class } k \text{ is distant from } j \end{cases} \tag{3.10}$$

with

$$0 \leq \alpha \leq \gamma \tag{3.11}$$

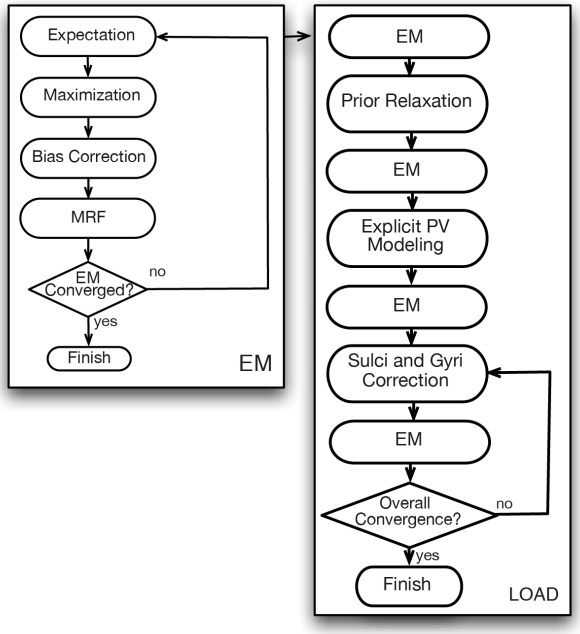


Figure 3.4: Algorithm flowchart.

where γ is a penalty factor for anatomically distant classes (*e.g.* eCSF and WM) and α is a penalty factor for anatomically neighbouring classes (*e.g.* eCSF and cGM). Under these assumptions, a bigger γ leads to a lower probability that two voxels with anatomically distant labels would be together and a bigger α would increase the sharpness of the transitions between neighbouring tissues, leading to more homogeneous but less detailed segmentations. The values for α and γ used in this section are 0.5 and 3 respectively.

Segmentation Refinement

The model described above is only based on global parameters. However, in some situations, due to lack of image contrast, INU, PVand noise, these global parameters are not enough to provide an accurate and topologically aware segmentation of fine structures. Three refinement levels were added to compensate for three main problems. First, a method was created to iteratively relax the constraints embedded within the prior information, compensating for problems in areas with high degree of natural anatomical or pathological variability. Second, an explicit modelling of PV was added and the MRF energy matrix was altered in order to incorporate the new classes. This refinement step obviates the need for an artificial threshold to separate pure and mixed voxels and allows different MRF behaviour between pure and PV corrupted areas. Finally, in order to add topological information to the segmentation and to increase the detail of the segmentation, a method to enhance the delineation of PV-corrupted grey matter folds is performed in an iterative manner until convergence. The algorithm's flowchart is shown in Fig.3.4

First Level: Prior Probability Relaxation

The EM algorithm is known to converge to a local maximum. In an ML approach, the prior probability drives the EM algorithm to a sensible solution, making it more robust to noise and INU. However, in

areas with high anatomical variability, prior driven ML approaches can lead to an erroneous solution because the prior probability for the expected class might be too close to 0 to allow the EM to converge to an anatomically feasible solution. It can also bias the segmentation towards the template, possibly overshadowing some anatomical differences. Here, a method where the prior probabilities are changed iteratively at each convergence of the EM algorithm is proposed. As our model parameters become closer to an anatomically feasible solution, one is able to locally relax our prior probability without reducing the robustness to noise, INU and PV. This is analogous to coarse-to-fine refinement of regularisation in image registration, for example the gradual reduction of prior influence over the outer iterations in DARTEL (Ashburner and Friston, 2009).

After the EM algorithm converges, the model parameters $\mathbf{\Phi}_y$ are closer to the optimal solution. However, due to the *a priori* spatial constraints, the segmentation of patients with different anatomical and structural characteristics might not converge to the correct solution. In order to relax these constraints and make the segmentation less dependant on these priors, one possible solution might be to smooth the priors and consequently smooth these constraints. However, because these relaxed priors would then be similar to uninformative priors, the problem would become similar to a Maximum Likelihood formulation, resulting in erroneous segmentations in patients with white matter hypo and hyper-intensities. Instead, similarly to Seghier et al. (2008), patient specific priors are generated using an ad hoc transformation over the posteriors. These updated atlases cannot be considered as priors in a strict mathematical sense as they are derived from the data, however they behave as such in this segmentation framework. The patient specific relaxed anatomical atlases are generated as a combination of the current estimates of the posteriors smoothed over anatomically neighbouring classes as described by

$$\pi_{ik} = \frac{p_{ik} + \sum_{j=1}^{K} \mathbf{H}(k,j)\tau_{ik}p_{ij}}{\sum_{l=1}^{K}\left(p_{il} + \sum_{j=1}^{K} \mathbf{H}(l,j)\tau_{ik}p_{ij}\right)} \tag{3.12}$$

with

$$\mathbf{H}(k,j) = \begin{cases} 0 & \text{if class } k \text{ is the same as } j \\ R_f & \text{if class } k \text{ is next to } j \\ 0 & \text{if class } k \text{ is distant from } j \end{cases} \tag{3.13}$$

and

$$\tau_{ik} = \frac{1}{1 + \mathcal{E}(p_{ik\,0.5})} \quad \text{and} \quad 0 \leq R_f \leq 1. \tag{3.14}$$

Here, τ_{ik} is inversely proportional to $\mathcal{E}(p_{ik})$, defined as the Euclidean distance from point i to the closest hard classified voxel where $p_{ik} > 0.5$. Thus τ_{ik} will be equal to 1 where $p_{ik} > 0.5$ and will have a decreasing value as the distance to the hard classified set increases. The parameter R_f controls the amount of prior probability sharing and $\mathbf{H}$ is a matrix containing the same anatomical neighbouring rules as the MRF.

Second Level: Explicit PV modelling

In PV segmentation, it is common to assume that if two tissues mix in a voxel, all mixing proportions are equally likely. The PV probability can be seen as a number of mixed Gaussians in between the two pure classes, corresponding to all the possible tissue proportions within a voxel (Van Leemput et al., 2003). Ruan et al. (2000) showed that, for brain imaging and for the signal-to-noise ratio and contrast-to-noise ratio levels of the current MRI systems, the density of all these PV Gaussian classes can be approximated by a single Gaussian with a small risk ($\alpha < 1$ for D'Agostino-Pearson normality test). Under this assumption, the values of p_{ik}, μ_k, σ_k are used to initialise an 8 class model, that considers the existence of the 6 original classes (now considered "pure") and 2 mixed classes $\{$WM, cGM, dGM, eCSF, iCSF, DU, WM/GM, GM/CSF$\}$. Even though every neighbouring class should have a mixed class in between, for the sake of computational complexity, the PV estimation is limited to the cortical layer. Using the same framework, the 8 classes are modelled as Gaussian mixtures on the log transformed data. The prior probability, average and variance for the 8 classes model are denoted as π^*_{ik}, μ^*_k and σ^*_k, where the superscript * is used to indicate that they belong to the 8 class model. While the 6 pure classes maintain their previous parameters, the initial mean, standard deviation and priors for the 2 mixed classes have to be estimated from the data. Under the assumption of Gaussian distributed classes on log-tranformed data, the initial mixed class Gaussian parameters can be approximated by a mixel distribution (Kitamoto and Takagi, 1999), with mean equal to the arithmetic weighted average of its composing class parameters weighted by each class's average fractional content. Thus,

$$\mu^*_{j/k} = \Gamma_{j/k}\mu_j + (1 - \Gamma_{j/k})\mu_k \tag{3.15}$$

where $\Gamma_{j/k}$ is the average of the fractional content (FC) between classes j and k for all voxels with $FC \in [0, 1]$. FC is defined as $\text{FC}_{JK} = (\mu_j - \bar{y_i})/(\mu_j - \mu_k)$ and $\bar{y_i} = y_i - \sum_j c_j\phi_j(x_i)$ is the image intensity corrected for INU. This is equivalent to calculating the average mixing vector $t = [\alpha, 1 - \alpha]$ in the model proposed by Van Leemput et al. (2003) for all the PV containing voxels and using that as a weighting factor. The initial value of the mixed class variance is estimated using the same mixel model. Assuming that the mixel variance is only dependent on his composing classes, the initial estimate of the mixed class variance can then be estimated by

$$(\sigma^2_{j/k})^* = \Gamma^2_{j/k}\,\sigma^2_j + (1 - \Gamma_{j/k})^2\,\sigma^2_k \tag{3.16}$$

Van Leemput et al. (2003) observed that the extra parameters or extra Gaussians that have to be introduced into the PV model hamper the segmentation robustness because no prior is available for the PV location. Our approach obviates this problem using information from the 6 class model to generate a patient specific spatial atlas, used as an ad hoc prior for the mixed classes. Due to the multiplicative nature of the probabilities, the mixed class prior is generated from the normalised geometric mean of its

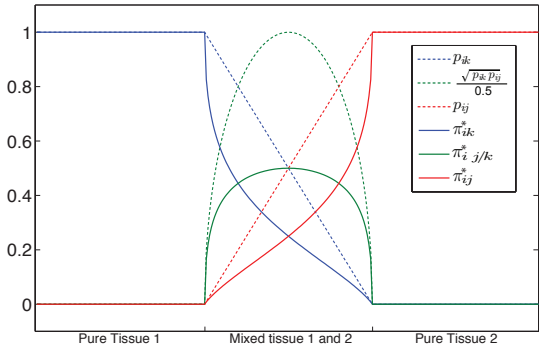


Figure 3.5: The mixed class prior (dashed green) is the normalised geometric mean of p_{ik} and p_{ij} (dashed blue and red respectively). The continuous lines represent their value after normalisation over all classes.

composing tissue distributions p_{ij} and p_{ij}, normalised over all classes.

$$\pi^*_{i(j/k)} = \frac{\sqrt{p_{ij}\,p_{ik}}}{0.5}\frac{1}{\Pi_i} \tag{3.17}$$

with Π_i as a normalisation constant over all classes (see figure 3.5). For the non mixed classes $\pi^*_{ik} = p_{ik}/\Pi_i$. The normalised geometric mean reflects how close p_{ik} and p_{ij} are from the situation where both composing tissues have equal proportions, having the value of 1 where $p_{ik} = p_{ij} = 0.5$ and 0 where either p_{ik} or p_{ij} are 0. One should bear in mind though, that $\pi^*_{i(j/k)}$ is not an estimation of the amount of PV, but just a geometrical transformation necessary to create priors for the mixed class. This new stage of the EM algorithm is initialised with $p_{ik} = \pi^*_{ik}$.

Third Level: MRF weighting for Deep Sulci and Gyri Delineation

As presented in Morris et al. (1996) and then discussed in Van Leemput et al. (2003) the MRF minimises the boundary length between tissues, discouraging classifications from accurately following the highly convoluted shape of the human cortex, resulting in incorrectly segmented structures such as deep sulci and gyri. Van Leemput et al. (2003) suggested that a nonstationary MRF model, with different parameters for uniform and convoluted regions, might be an interesting solution to the the MRF problem. This is exactly the problem that this section tries to solve. (Fischl et al., 2002) used a spatially varying MRF prior in order to increase the local label neighbourhood adaptiveness and robustness. Even with non empirical estimation of warp regularisation parameters (Yeo et al., 2008), the creation of sharp priors for this purpose is difficult due to the highly variable sulcal and gyral location. Thus, this method still does not optimally address the MRF bias-variance tradeoff. Instead, a modified version of the current posterior estimates is used in order to generate a patient specific sulci and gyri atlas and use this information as an MRF strength weighting. Even though it is an ad hoc modification, it enables a robust and sharp localisation of these structures, improving the delineation of the cortical folds. In a similar way to (Acosta et al., 2008) and (Han et al., 2004), the information derived from a distance transform is used to estimate the location of deep sulci and gyri and change the priors and the strength of the MRF only in those locations. Cost functions based on the norm of the gradient of the Euclidean distance transform,

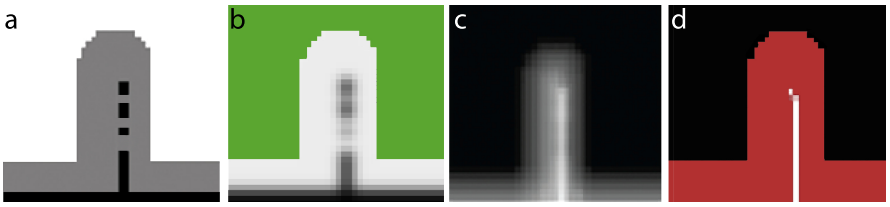


Figure 3.6: Sulci localisation using the proposed metric. (a) Current binary segmentation, (b) hard segmented set in green with the respective speed function s_j in grey levels, (c) geodesic distance (time of arrival), (d) the phantom in red overlaid with the detected sulci location in white

like the one used in (Acosta et al., 2008), have several drawbacks: Using a Euclidean based distance implicitly assumes that both banks of the sulci or gyri have the same thickness which is frequently not true; the metric is non informative with regards to the current PV estimates; it overlooks the fact that the norm of the gradient can be zero in both local maxima or minima, whereas the areas of interest should always be in local maxima. The cost function proposed by (Han et al., 2004) uses the estimated segmentation to add information about the sulci position, however it still suffers from the same mathematical drawbacks as it is also only based on the gradient of the distance. In order to improve on these limitations, a previously published method (Cardoso et al., 2010) was used to detect the sulci and gyri location.

The assumption that both banks of the sulci and gyri have the same thickness can be removed by using the segmentation probabilities as a speed function for an evolving level set. Figure 3.6 (a) shows the current hard classification of GM, WM and CSF. In (b), the green area is the initial estimate of the level set, initialised from the current hard WM segmentation. This green surface evolves with a speed inversely proportional to the WM probability. Figure 3.6 (c) shows the resulting geodesic distance (time of arrival) for the evolving front. Both sides of the evolving front will stop as they meet, thereby defining the position of the sulci. These locations are then fed-back into the segmentation framework by locally weighting the MRF and changing the priors (Cardoso et al., 2010). The same process evolving from the eCSF twards the WM will detect the WM stalks.

The functions ω_i^{gyri}, ω_i^{sulci}, used to weight the MRF, are defined as follows:

$$\omega_i^{\text{gyri}} = \mathcal{H}\big(-\nabla.\nabla\mathcal{G}_i(h_{\text{WM}}, \tfrac{\xi}{(\xi+p_{\text{CSF}})})\big)\mathcal{H}\big((1-||\nabla\mathcal{G}_i(h_{\text{WM}}, \tfrac{\xi}{(\xi+p_{\text{CSF}})})||)\big) \tag{3.18}$$

$$\omega_i^{\text{sulci}} = \mathcal{H}\big(-\nabla.\nabla\mathcal{G}_i(h_{\text{CSF}}, \tfrac{\xi}{(\xi+p_{\text{WM}})})\big)\mathcal{H}\big((1-||\nabla\mathcal{G}_i(h_{\text{CSF}}, \tfrac{\xi}{(\xi+p_{\text{WM}})})||)\big) \tag{3.19}$$

where $\nabla.\nabla$ is the Laplacian operator, $\mathcal{G}_i(h_k, s_j)$ is the geodesic distance transform (computed using the Eikonal equation $\mid \nabla\mathcal{G} \mid = s_j$) from the hard segmentation set $h_k = p_{ik} > 0.5$, $s_j = \xi/(\xi + p_j)$ is a speed function, $\xi = 0.05$ and $\mathcal{H}$ as a limiting function defined as,

$$\mathcal{H}(x) = \begin{cases} 1 & x \geq 1 \\ x & 1 > x > 0 \\ 0 & x \leq 0 \end{cases} \tag{3.20}$$

The limiting function is necessary due to the behaviour of the first and second derivatives of $\mathcal{G}_i$ in areas where the speed function is close to zero. It also clips the negative component of $\nabla.\nabla\mathcal{G}$, removing the influence of the local minima in the overall cost function. Furthermore, the clipping effect leads to an ω function that is sharp (close to one voxel thick) enforcing a minimum opening. This was done by design since one would expect a sulcus or gyrus of more than two voxels thick to be already correctly classified. The constant ξ is set to 10^{-6}. An example of $\mathcal{G}$ and ω is shown in Fig.3.7.

The main advantage of a divergence based metric is the ability to distinguish between local maxima and minima, improving the robustness of the sulci and gyri detection. In order to introduce local adaptivity of the MRF, a local weighting function is incorporated in Equation 3.8 by making β_i a spatially-varying value

$$\beta_i = (1 - \omega_i^{\text{sulci}})\,(1 - \omega_i^{\text{gyri}}) \tag{3.21}$$

Normally β_i corresponds to the overall MRF strength, however, in this case, the overall MRF strength is directly introduced into the α and γ penalty factors and β_i only acts as a local weighting. The values of ω^{sulci} and ω^{gyri} vary between [0,1] and have a value of 1 near the centre of the sulci and the centre of the gyri respectively. In a similar way, the value of β_i lies between [0,1] and has a value of 0 near the centre of the sulci and gyri.

The functions ω_i^{sulci} and ω_i^{gyri} are also used to iteratively change π_{ik} to give more prior probability to the respective classes in areas where deep sulci and gyri should exist.

For classes WM/GM, GM and GM/CSF, π_{ik} is updated as

$$\pi^*_{i\,(\text{WM/GM})} = p_{i\,(\text{WM/GM})} + (\omega_i^{\text{gyri}} p_{i\text{GM}}) \tag{3.22}$$

$$\pi^*_{i\,(\text{GM})} = p_{i\,\text{GM}}\,\beta_i \tag{3.23}$$

$$\pi^*_{i\,(\text{GM/CSF})} = p_{i\,(\text{GM/CSF})} + (\omega_i^{\text{sulci}} p_{i\text{GM}}) \tag{3.24}$$

The values of π_{ik} are then normalized in order to sum to one. Both the MRF's β_i and the priors π_i^* are updated every time the EM converges, and a new EM starts. The algorithm finishes when the ratio of likelihood change is less than a predefined ε, here set to 10^{-3}.

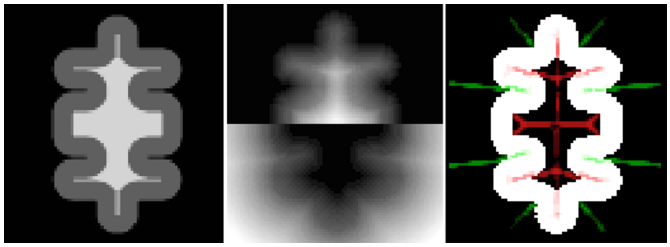


Figure 3.7: Sulci and gyri enhancement: (Left) Expected segmentation; (Centre) $\mathcal{G}(h_{\text{CSF}}, s_{\text{WM}})$ and $\mathcal{G}(h_{\text{WM}}, s_{\text{CSF}})$ on the top and bottom respectively; (Right) ω_i^{sulci} and ω_i^{gyri} in green and red respectively.

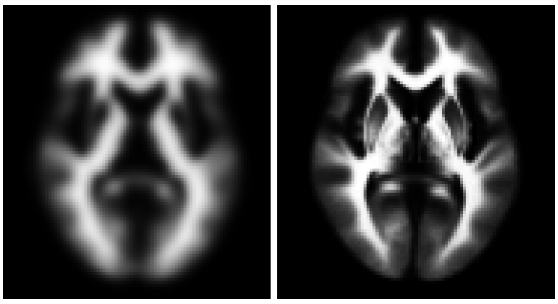


Figure 3.8: (Left) The MNI305 atlas and (Right) the ICBM452.

3.1.3 Experiments and Results

In this section, the proposed cortical segmentation algorithm was evaluated in terms of its independent parts and its overall performance. The first two experiments are intended to show the contribution of each component to segmentation performance. The proposed method was then evaluated globally against synthetic and clinical data in order to access the accuracy of the PV estimation, segmentation overlap and group separation and additionally, the method was compared to three state of the art methods: FANTASM [(Pham, 2002b)], SPM8 [(Ashburner and Friston, 2005)] and FAST [(Zhang et al., 2001)]. The first method is a fuzzy c-means based segmentation with bias field optimisation and noise tolerance. The second method is an EM based iterative segmentation/registration algorithm with bias correction and the last method is an EM based segmentation, specifically chosen because it uses an MRF to add spatial consistency. In all statistical tests the significance level was set to $p < 10^{-3}$. Unless mentioned otherwise, the relaxation fraction $R_f = 1$.

Atlas dependency study

A segmentation algorithm that is fully independent from the chosen atlas is expected to produce the same result when segmenting a dataset with two different atlases. However, the use of different atlases changes the prior probability and thus the final segmentation results. In order to evaluate the segmentation dependency on the atlases and the effect of the prior relaxation, a subset of 40 subjects, 20 patients diagnosed with AD and 20 age- and gender-matched controls were selected from the ADNI database. These datasets were segmented using two different anatomical atlases and 4 different relaxation factors R_f between 0 and 1, leading to 320 different segmentations. The two different atlases were the ICBM452 and the smoothed version (8mm FWHM) of the MNI305 (Evans et al., 1993). The ICBM452 was created by non-rigidly registering and averaging 452 normal MRI scans while the MNI305 was created by affinely registering and spatially smoothing 305 normal MRI scans. Both atlases are representative of a normal population, with the main difference being the registration method and the amount of blurring used to create them (see Figure 3.8).

For each dataset and relaxation factor, a fuzzy Dice score (Crum et al., 2006) was calculated between the cortical GM segmentations obtained using the two atlases. The fuzzy Dice score assesses the overlap and the PV differences between the segmentations without the need for a threshold value. The results are shown in Figure 3.9. When the prior relaxation is applied to the control population there is almost zero

difference in the Dice score average and just a small decrease in the standard deviation. However, when the prior relaxation is applied to an AD population, there is an upward trend in the median Dice score and a reduction in the interquartile difference when the relaxation factor is increased, with the median Dice score going from 0.906 to 0.924.

Thickness measurement evaluation

Voxel-based cortical thickness measurements are critically dependent on the quality of the segmentation and its topology. As there is no ground truth, a digital phantom was used in order to assess the accuracy and precision of thickness measurements.

A very high resolution digital phantom containing finger and sheet like collapsed sulci and gyri was created, simulating the complex and convoluted structure of the cortex. The phantom's white matter is homotopic to a ball and the cortical layer has a Euclidean thickness of 8 mm between the inner and outer surface. The phantom was created on an 0.25 mm isotropic image resulting in $600 \times 600 \times 1000$ voxels. The thickness of the high resolution phantom was calculated using a Laplace equation based method (Acosta et al., 2009). Due to the curved nature of the Laplace equation's streamline, the resulting ground truth mean (standard deviation) thickness was 8.13 (0.15) mm. The phantom was then Fourier-resampled to reduce the resolution by a factor of 5 in each dimension. Two levels of complex Gaussian noise were also added in the Fourier domain, resulting in two low resolution Rician noise corrupted phantoms. To obtain artificial anatomical priors for the segmentation step, the ground truth segmented images were Gaussian filtered ($\sigma = 4$ mm) to simulate the anatomical variability. The thickness was then measured on the segmented low resolution phantoms using a Laplace equation based method with a Eulerian-Lagrangian approach as described in (Acosta et al., 2009).

The results are shown in Fig.3.10 and Table 3.1. When compared to the ground truth, the proposed method yields a difference in the average thickness of 0.14 mm and 0.48 mm for the low and high noise phantoms respectively. The standard ML approach with the MRF but without the proposed improvements yields a difference in average thickness of 4.74 mm and 4.36 mm for the low and high noise phantoms respectively. Finally, the standard ML approach without either the MRF or the proposed improvements yields a difference in average thickness of 3.98 mm and 1.22 mm for the low and high noise

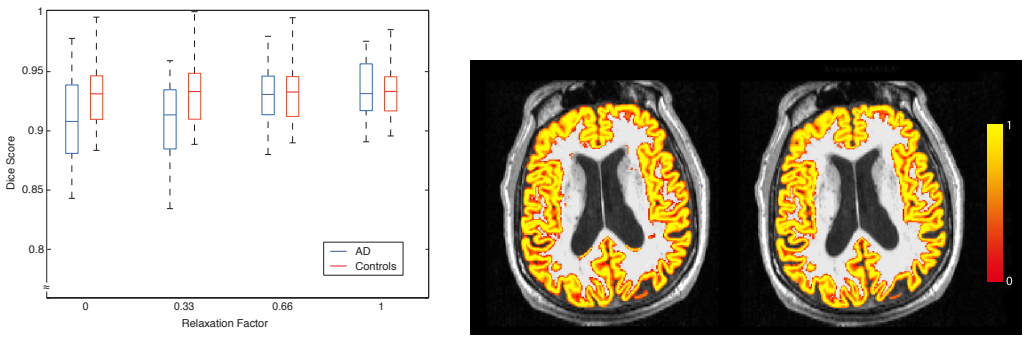


Figure 3.9: (Left) The fuzzy Dice scores between the cortical GM segmentations using different atlas and relaxation factors. Segmentation example with Relaxation Factor $= 0$ and Relaxation Factor $= 1$ (Right). Notice the improved segmentation results in the ventricle area.

Table 3.1: Table contains the thickness average and standard deviation for the three methods and two levels of noise.

	Low noise mean (std) mm	High noise mean (std) mm
ML without MRF	12.11 (2.55)	9.35 (3.10)
ML with MRF	12.87 (2.98)	12.48 (2.82)
Proposed Method	8.27 (0.32)	8.61 (0.91)

phantoms respectively.

Segmentation evaluation

20 datasets were downloaded from the BrainWeb (`http://www.bic.mni.mcgill.ca/brainweb`) MR image simulator. Each dataset contained a simulated T1-weighted image and a corresponding ground truth grey matter probabilistic atlas. The simulated data was generated using a spoiled FLASH sequence with TR $=$ 22 ms, TE $=$ 9.2 ms, $\alpha = 30°$ and 1-mm isotropic voxel size with simulated 3% noise and 20% INU (Aubert-Broche et al., 2006). The 20 simulated images were segmented using the proposed method, SPM8, FAST and FANTASM, each one resulting in either a PV segmentation or its fuzzy c-means equivalent. The analysis focuses on the GM class as most of the differences between the methods will be in the cortical area

For each segmentation, a normalised cumulative histogram of the absolute difference between the segmentation and the ground truth was calculated. Figure 3.11 (a) shows the mean and standard deviation as error bars for the 20 datasets. The proposed method results in 94% of voxels having an absolute difference of less than 0.1 compared to 87% for FAST, 84% for SPM8 and 80% for FANTASM.

Figure 3.11 also shows p-values calculated using a two-tailed unequal-variance two-group t-tests between the normalised absolute difference histogram values of our method and each of the other two methods. The proposed method achieves significantly better PV estimation than FAST, SPM8 and FAN-

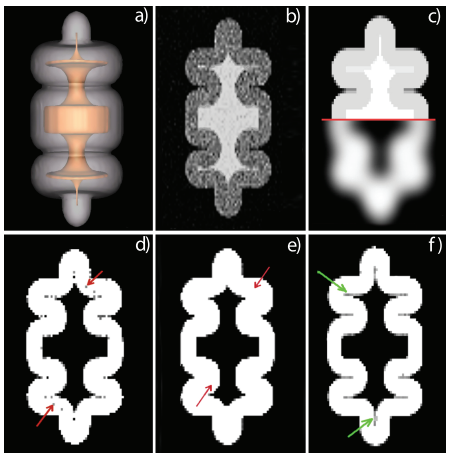
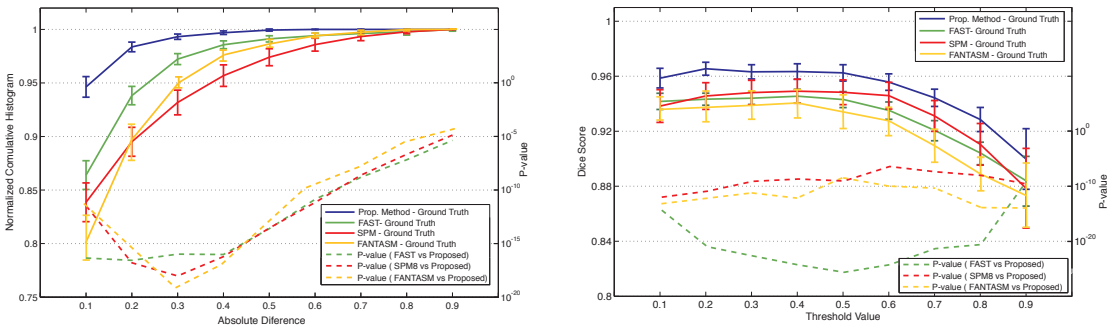


Figure 3.10: Phantom segmentation and thickness results: a) 3D model of the phantom, b) High noise phantom, c) True labels and GM prior used, d) ML without MRF, e) ML with MRF, f) Proposed method. The red arrows point to the presence of noise and lack of detail causing wrong thickness measurements. The green arrows point to the detected deep gyri.



a b

Figure 3.11: (a) Normalised cumulative histogram of the absolute difference between the segmentation and the ground truth; (b) Dice score between the segmentation and the ground truth at several threshold values.

TASM for all threshold values.

To evaluate the quality of the binarised and PV segmentations, the binary and fuzzy Dice scores (Zijdenbos et al., 1994; Crum et al., 2006) were calculated between the segmentations and the ground truth. The binary Dice score was calculated in order to understand the behaviour of the the overlap metric with regards to the threshold level. Here, the binary Dice score was estimated at several PV thresholds and two-tailed unequal-variance two-group t-tests were calculated between the proposed method, FAST, SPM and FANTASM. Figure 3.11 (b) shows the average Dice score and standard deviation as error bars for the 20 datasets and the results of the statistical test. For all threshold values, the proposed method achieved significantly higher average Dice scores than FAST, SPM and FANTASM. The fuzzy Dice score was calculated in order to give an overall measure of unthresholded segmentation alignment. Here, the average fuzzy Dice score for the 20 datasets was 0.959, 0.941, 0.929 and 0.927 for the proposed method, FAST, SPM and FANTASM respectively.

ADNI Data Study

To further investigate if the proposed refinements are useful when extracting measurements from the segmentation, cortical thickness was calculated on ADNI data in order to evaluate group separation between controls and Alzheimer's Disease (AD) diagnosed patients. This metric was chosen because it is dependent on both the accuracy and the topology of the segmentation. A subset of the ADNI database was used in this study. From the full database, 88 AD diagnosed patients and 82 age- and gender-matched controls were selected, with T1-weighted volumetric images acquired on 1.5 T units using 3D MPRAGE or equivalent protocols with varying resolutions (typically 1.25 x 1.25 x 1.2 mm).

All 170 datasets were segmented using the proposed method and the two best methods with regards to the fuzzy Dice score from the previous section - SPM8's standard unified segmentation and FAST. In order to test only the improvements of the proposed methodology, the cortical thickness was calculated using a Laplace equation based algorithm (Acosta et al., 2009) and not the algorithm proposed in Section 5.1. This method requires the user to select a threshold to classify a voxel as pure (normally 0.95) in order to solve the Laplace equation. This threshold in normally set high and not at the optimum Dice

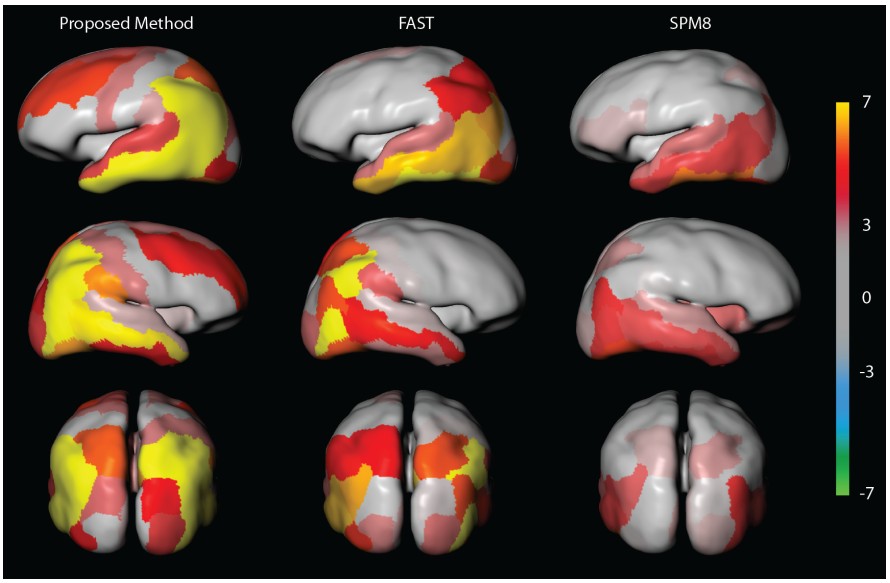


Figure 3.12: Statistical significance of cortical thickness between AD patients and controls: Colour coded p-values are represented in logarithmic scale with positive and negative values associated with thinning and thickening respectively.

score in order increase the level of detail in the obscured sulci and gyri area, resulting in less biased thickness measurements. As both FAST and the proposed method use an MRF to add spatial consistency, a voxel was considered pure when $p_{GM} = 1$. However, for SPM8, a voxel was considered pure for $p_{GM} > 0.95$ to compensate for the lack of MRF. The same transformation used to map the priors to the individual subjects was used to propagate the AAL template (Tzourio-Mazoyer et al., 2002), and the average thickness at the central Laplacial isoline was calculated for 52 AAL cortical regions. Two-tailed unequal-variance two-group t-tests between patients and controls over each AAL region were calculated. In order to visualise the results (Fig. 3.12), log transformed p-values were propagated back to the corresponding areas on a normal population smoothed 3D model with positive and negative values associated with thinning and thickening respectively. The p-values were thresholded at $p = 10^{-3}$. The expected areas affected in AD patients are the middle and inferior temporal, superior and inferior parietal and middle frontal gyrus bilaterally. Using the proposed method as segmentation, all of these areas show statistically significant differences in thickness with $p < 10^{-5}$ in the middle temporal and parietal regions and $p < 10^{-3}$ in the middle frontal gyrus region. Although most of the same expected areas are statistically significant when using FAST's segmentation, the middle frontal gyrus area does not show statistically significant differences. Also, only the left middle and inferior temporal regions and right parietal region show statistically significant differences in thickness with $p < 10^{-5}$ leading to a noticeable lack of symmetry between hemispheres. Using SPM, there is an overall decrease of statistical significance throughout the brain, with only the middle and inferior temporal areas above the $p < 10^{-3}$ threshold.

Computation Time

The total computation time is in line with current state of the art segmentation methods. The segmentation step takes on average less than 2 minutes, with an overhead of less than 3 min for the registration of the priors since the registration is fairly broad, resulting in an average total time below 6 minutes per dataset.

3.1.4 Discussion

This section (3.1) of the thesis presents a new segmentation method specifically designed for the cerebral cortex. The robustness and accuracy of the segmentation and PV estimation and also the ability to directly use the segmentation for cortical thickness estimation on synthetic and real data was evaluated.

In section 3.1.3, a study testing for atlas independence was performed on real data from the ADNI database in order to evaluate the efficacy of the prior relaxation. When segmenting the datasets using two normal population atlases, an algorithm that is less dependent on the prior probability would produce two closely matching segmentations. As expected, the results show that when priors derived from a control population are applied to a control group, there is no change in the average dice score, since the atlas is representative of that specific population. However, when a control population atlas is applied to an AD population, an increase of the relaxation factor has a positive effect on the segmentation overlap. Although the difference is not significant, there is an upward trend on the average and a decrease on the standard deviation of the Dice score distributions. This shows that after prior relaxation, the segmentations become more similar, and thus, less dependent on the priors. Visual assessment shows a noticeably better segmentation in the ventricle area of the AD patients, mainly when the ventricles are expanded (see Figure 3.9). This is caused by the spatial ambiguity when the ventricle edge is close to the cortical GM. A higher relaxation factor also produces a visually better temporal lobe segmentation when these are highly atrophied. Overall, the extra knowledge introduced in the prior relaxation step by the neighbouring tissue structure leads to reduced bias, resulting in less ambiguity regarding miss-segmented areas due to different anatomy.

A second experiment showed that the proposed improvements can help to accurately extract meaningful thickness measurements from the segmentation. Using a digital phantom created specifically for this purpose, the average thickness was measured with the proposed method, without the refinement steps (MAP with MRF), and just using the intensity component of the model (MAP without MRF). The results are displayed in Table 3.1. Consistent results were found for both low and high noise cases. An unweighted MRF caused an overestimation of the thickness and standard deviation due to the lack of detail in highly convoluted and PV corrupted areas. Without any type of MRF, the thickness measurements are much more prone to noise, leading to a number of short paths to mis-segmented voxels and consequently an artificial increase of the standard deviation of the measurement. Oddly, when the noise level is high, the presence of short paths combined with the lack of detail leads to a more accurate estimate of the average thickness. However, because the standard deviation is much higher than expected, this measurement lacks precision.

In section 3.1.3, the Dice score and PV estimation accuracy were evaluated using BrainWeb data.

The proposed method and FAST both showed higher PV estimation accuracy than SPM8 and FANTASM. This is most probably due to the MRF smoothing properties that make the PV estimation more robust. Also, the MRF will ensure a more robust assignment of voxels surrounded by only one tissue class, thus making the posterior probabilities rmore closely resemble PV fractions. The small Dice score improvment of the proposed method can be explained by the adaptive nature of the MRF in areas close to sulci and gyri, increasing the level of detail whilst maintaining robustness to noise. On the other hand, due to the lack of adaptivity in FAST's MRF, some of the details are lost, leading to worse PV estimation when compared to the proposed method. SPM8 underperforms both FAST and the proposed method with regards to PV estimation accuracy. One can speculate that for cortical segmentation specifically, the advantages of having an iterative segmentation/registration procedure may not compensate for the lack of MRF. Finally, even though FANTASM is tolerant to noise, it does not model noise implicitly. This might explain the small underperformance with regards to Dice score of FANTASM over the other methods for low PV differences. The difference between FANTASM and the proposed method becomes smaller for difference values above 0.3.

The proposed method achieved significantly higher Dice scores when compared to FAST, SPM and FANTASM. I believe the improved overlap between structures is probably due to the enhanced delineation of the sulci and gyri and implicit PV modelling. Also because these improvements are iteratively fed back into the segmentation, there is a gradual reduction of the PV related parameter bias. One might also conclude that SPM outperforms FAST in terms of Dice score due to the iterative segmentation/registration procedure, improving the overlap of the segmented structures. Another explanation might be the lack of spatial adaptiveness in FAST's MRF, as the MRF tends to minimize the boundary length between tissues which discourages classifications from accurately following the highly convoluted shape of the human cortex. For the proposed method, this problem is reduced as the MRF is spatialy adaptive.

In the fourth experiment, using ADNI data, the three segmentation methods were compared in terms of group separation between control subjects and Alzheimer's Disease (AD) diagnosed patients. Using the proposed segmentation, statistically significant clinically-expected pattern of difference in cortical thickness between AD patients and controls was detected. Although most of the same expected areas are also statistically significant when using FAST's segmentation, there is a less symmetric pattern of atrophy and some of the expected areas (*i.e.* right and left middle frontal gyrus) don't achieve statistical significance. This is probably caused by the lack of detail due to the use of a stationary MRF. When using SPM, there is a noticeable overall decrease of statistical significance throughout the brain, with only the middle and inferior temporal areas achieving statistical significance. This is again caused by the lack of detail, mostly due to the need for an artificial threshold to separate pure from non-pure voxels. This shows how important the presence of an MRF is when segmenting the cortex. Throughout the literature, the vast majority of clinical studies have been carried out using surface-based cortical thickness techniques (Lerch et al., 2005; Du et al., 2007; Lehmann et al., 2009; Rosas et al., 2008; Nesvåg et al., 2008; Salat et al., 2004) with a few using voxel-based techniques (Querbes et al., 2009). Both methods depend

on the segmentation step; however, for surface-based techniques, the segmentation is only used as an initialisation for a surface mesh. The mesh is typically deformed to fit the cortical GM/WM boundary and expanded outwards to the GM/CSF boundary. This gives surface-based methods sub-voxel accuracy and robustness to noise. However, due to smoothness and topology constraints, it is difficult to correctly fit the surface to very complex shapes thus requiring laborious manual corrections. Additionally, the implicit surface modelling can lead to bias in the thickness measurements (MacDonald et al., 2000; Kim et al., 2005). Conversely, voxel-based techniques can potentially cope with any topology or shape because they work on the 3D voxel grid. However, these techniques were never specifically tailored for the highly convoluted shape of the cortex. The proposed segmentation method improves the quality and topology of the cortical segmentation and enhances the detection of PV corrupted sulci and gyri, enabling the direct use of the segmentation for cortical thickness as opposed to requiring post-processing techniques (Hutton et al., 2008; Lohmann et al., 2003; Acosta et al., 2009).

3.1.5 Conclusions

This section of the thesis presents a segmentation algorithm tailored for applications such as cortical thickness estimation. The main contributions of this section lies in three refinement steps: a method that iteratively relaxes and modifies the prior information in an anatomically coherent way to reduce the bias towards the priors; the explicit modelling of the PV effect and the adaptation of the MRF energy to reflect the inclusion of these new classes; the introduction of a new distance based cost function to add information about the location of PV corrupted grey matter folds and integrated that information into the segmentation framework.

The method achieves more accurate and precise delineation of collapsed grey matter folds without losing robustness to noise and intensity inhomogeneity. Even though some of these refinement steps can be considered as *ad-hoc*, they are integrated within a single framework. Quantitative analysis on 20 BrainWeb datasets showed statistically significant improvements in the accuracy of the PV estimation and in the Dice score when compared to three state of the art techniques. Cortical thickness measurements on a new digital phantom demonstrated improvements in the accuracy and robustness of the thickness calculation using the proposed method, when compared to other methods. Results on ADNI data showed clinically-expected patterns of cortical thinning between AD patients and controls using the proposed method, with highly significant group differences in several expected regions.

3.2 AdaPT: an adaptive preterm segmentation algorithm for neonatal brain MRI

Babies born prematurely are at increased risk of adverse neurodevelopmental outcomes. Recent advances suggest that measurement of brain volumes can help in defining biomarkers for neurodevelopmental outcome. These techniques rely on an accurate segmentation of the MRI data. However, due to lack of contrast, PV effect, the existence of both hypo- and hyper-intensities and significant natural and pathological anatomical variability, the segmentation of neonatal brain MRI is challenging. This section presents a pipeline for image segmentation that uses a novel multi-model Maximum a posteriori Expectation Maximisation (MAP-EM) segmentation algorithm with a prior over both intensities and the tissue proportions, a B0 inhomogeneity correction, a spatial homogeneity term through the use of a Markov Random Field, an adaptive technique that enables the segmentation of images with high anatomical disparity from a normal population and implicit PV modeling.

3.2.1 Derived publications and methodological sumary

- **Cardoso, M. J.**, Melbourne, A., Kendall, G.S., Modat, M., Robertson, N.J., Marlow, N., Ourselin, S.: AdaPT: an adaptive preterm segmentation algorithm for neonatal brain MRI. (2012) Neuroimage

- Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Ourselin, S., Radial structure in the preterm cortex; persistence of the preterm phenotype at term equivalent age? (2012) MICCAI

- Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Ourselin, S., NeoBrainS12 Challenge: Adaptive neonatal MRI brain segmentation with myelinated white matter class and automated extraction of ventricles I-IV (2012) MICCAI NeoBrainS

- Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Ourselin, S., A cortical surface analysis of very preterm infants on term-equivalent age MRI (2012) MICCAI PAPI

- **Cardoso, M. J**., Melbourne A., Kendall G. S., Modat M, Hagmann C. F., Robertson N. J., Marlow N., Ourselin S.; Adaptive Neonate Brain Segmentation. (2011) MICCAI

- Melbourne A., Kendall, G.S., **Cardoso, M.J.**, Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Analysing the cortical folding pattern of very preterm neonates scanned at term-equivalent age: Correlations with diffusion tensor tractography (2012) ISMRM

- Melbourne A., Kendall, G.S., **Cardoso, M.J.**, Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Analysing the cortical folding pattern of very preterm neonates scanned at term-equivalent age: Correlations with diffusion tensor tractography (2012) ISMRM

- **Cardoso, M.J.**, Melbourne A., Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Outlier Rejection for Adaptive Neonatal Segmentation (2012)

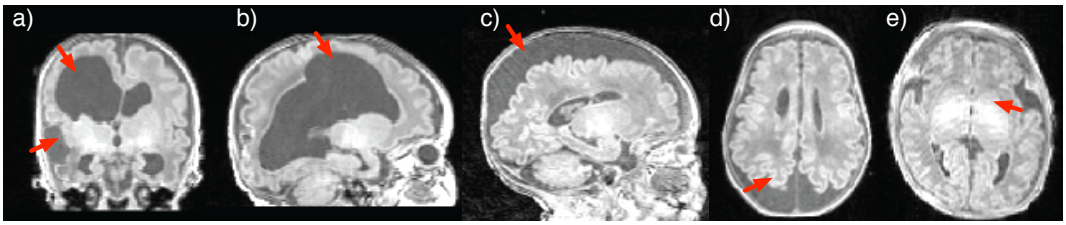


Figure 3.13: Red arrows pointing to the problematic areas. Pathological areas that differ from the normal population in patients with (a) and (b) ventriculomegaly, (c) and (d) excessive sub-arachnoid CSF. (e) Movement artefacts and hyper-intensities.

ISMRM

- **Cardoso, M.J.**, Melbourne A., Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Adaptive neonatal brain segmentation: application to ventriculomegaly and excess extra-axial cerebral-spinal fluid. (2011) ESPR

- Melbourne A., **Cardoso, M.J.**, Kendall, G.S., Hagmann, C.F., Bainbridge, A., Marlow, N., Robertson, N. J., Ourselin, S., Automated Analysis of the Preterm Neonatal Cortex at Term Equivalent Age and Correlation with Cognitive Outcome at 1 Year Corrected Age (2011) ESPR

In this section, a new neonate specific segmentation pipeline is presented. This framework incorporates a novel *Maximum a Posteriori* Expectation-Maximization (MAP-EM) based probabilistic segmentation technique that includes INU correction, spatial dependence via a Markov Random Field (MRF) and PV containing voxels. The pipeline also iteratively relaxes normal population priors, thus enabling their adaptation to pathological cases. The segmentation results are then used to iteratively update the registration from the template space to the subject, improving the prior alignment, the segmentation and the skull-stripping.

3.2.2 Methods

The segmentation procedure combines a number of steps which are now outlined in detail: the brain extraction procedure, expectation maximisation framework (incorporating a novel prior on the expected tissue intensities), the spatial regularisation process, the prior relaxation strategy and finally the neonate-specific PV correction strategy. In order to simplify the segmentation process, the neonatal brain volumes must first be extracted from the full image. Due to the white/grey matter contrast inversion, low signal-to-noise ratio and high anatomical variability, publicly available brain extraction algorithms (Smith, 2002; Segonne et al., 2004) might not achieve acceptable results. For the neonatal brain segmentation, we use a multi-atlas based segmentation propagation scheme. The brains of 15 neonates, including both normal and pathological subjects (infants with ventriculomegaly or extra-axial CSF) were selected from the entire cohort (section 3.2.3) and were first manually segmented using the ITK-SNAP software (Yushkevich et al., 2006). These manual brain masks, together with the associated T1-weighted images, represent the template database. In order to extract the brain of a new subject, each of the T1-weighted

images from the template database is aligned with the new image using an affine registration (Ourselin et al., 2000) followed by a free-form non-rigid registration algorithm (Modat et al., 2010). The manual segmentation of each atlas is then transformed using the deformation determined during the registration step. Local fusion of the propagated segmentations then forms a single consensus segmentation estimate using a modified version of the Locally Normalised Cross Correlation (LNCC) STAPLE framework as described in Cardoso et al. (2011c). Following the brain extraction routine, subject specific priors are generated by non-rigid alignment of the atlas tissue intensity template to the subject space (Kuklisova-Murgasova et al., 2011) and subsequent propagation of the associated tissue class priors. In addition to grey matter, white matter and cerebrospinal fluid classes, this atlas additionally defines sub-classes of the deep grey matter, cerebellum and brainstem.

MAP Expectation-Maximization Segmentation

After brain extraction and tissue class prior alignment, the segmentation proceeds using a *maximum a priori* expectation maximisation algorithm that we now describe in detail. Although in this work we apply the procedure to unimodal data, we present the method in general form so that it may be applied directly to (registered) multimodal data. This method will expand and generalise the single modality framework presented in section 3.1.

Assuming either single modality or coregistered multimodal datasets, let $i \in \{1, 2, \cdots, n\}$ index the n voxels of an image domain, with its intensities forming a feature vector $\mathbf{y}_i \in \mathbb{R}^m$. For K tissue types, let $\mathbf{z}_i$ denote the tissue type of voxel i. Thus, $\mathbf{z}_i = \mathbf{e}_k$ for some k, $1 \leq k \leq K$ where $\mathbf{e}_k$ is a unit vector with the kth component equal to one and all the other components equal to zero. Let each tissue type be described as having multivariate normally distributed intensities with mean vector $\boldsymbol{\mu}_k \in \mathbb{R}^m$ and respective covariance matrix $\boldsymbol{\Sigma}_k \in \mathbb{R}^{m \times m}$ grouped in $\boldsymbol{\theta}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$. In order to correct for MRI field inhomogeneity, we assume the presence of a multiplicative bias field and model INU as a linear combination $\boldsymbol{c}^T\phi$ of J smoothly varying basis functions $\phi(x) = \{\phi(x)_1, ..., \phi(x)_J\}$, where x denotes the spatial position and $\mathbf{c} = \{\mathbf{c}_1, ..., \mathbf{c}_m\}^T$ with $\mathbf{c}_m$ as a vector of coefficients for channel m. For mathematical convenience and similarly to Wells III et al. (1996) and Zhang et al. (2001), we assume that the intensity of the voxels that belong to class k are normally distributed after log transformation. This log transformation of the data makes the multiplicative bias field additive, enabling the existence of a linear least square solution for the coefficient optimisation.

Let $\boldsymbol{\Phi}_y = \{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,...,\boldsymbol{\theta}_K,c\}$ represent the overall model parameters. Assuming an *a priori* distribution over these parameters, the problem can be formalised as a *Maximum a Posteriori* (MAP) estimation of $\boldsymbol{\Phi}_y$ by

$$\hat{\Phi}_y = \arg\max_{\Phi} f(\mathbf{y} \mid \boldsymbol{\Phi}_y) g(\boldsymbol{\Phi}_y) \tag{3.25}$$

with $f(y \mid \boldsymbol{\Phi}_y) = \prod_i \sum_k f(\mathbf{y}_i \mid \mathbf{z}_i = \mathbf{e}_k, \boldsymbol{\Phi}_y) f(\mathbf{z}_i = \mathbf{e}_k)$ and $g(\boldsymbol{\Phi}_y)$ as a semi-conjugate Gaussian prior over the mean. The addition of these priors is a novel step, included here to improve the algorithms robustness to variations in neonatal image contrast. The priors act to constrain the space of solutions of the mean, increasing algorithm robustness in complex segmentation cases. Here, the semi-conjugate

prior $g(\mathbf{\Phi}_y)$ over the parameter $\boldsymbol{\mu}_k$ is defined as

$$g(\boldsymbol{\mu}_k \mid \theta_k^*) = \frac{1}{(2\pi)^{\frac{m}{2}} \mid \boldsymbol{\Sigma}_k^* \mid^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^*)^T (\boldsymbol{\Sigma}_k^*)^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^*)} \tag{3.26}$$

with $\mid \boldsymbol{\Sigma}_k^* \mid$ as the determinant of $\boldsymbol{\Sigma}_k^*$ and $\theta_k^* = \{\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*\}$ as the priors over the parameter $\boldsymbol{\mu}_k$. These parameters are obtained *a priori* by manually sampling for each tissue class from the set of 15 intensity normalised images used for the brain extraction step above. Intensity normalisation of each dataset is done by calculating a simple robust max-min (assuming 2% outliers) after masking. Then, for each subject, the mean intensity for each tissue class is calculated by manually sampling the intensity from several 3x3x3 blocks. From these samples, a population-specific distribution of the patch mean intensity for each tissue is obtained. Finally, given a new image (normalised using the same robust max-min approach), the population tissue distributions are used as priors over the Gaussian model parameters. On multimodal images, the process would be carried out separately for each modality if $\boldsymbol{\Sigma}_k^*$ is assumed to be diagonal. The inclusion of this extra term adds extra knowledge about the class specific distribution of the parameter μ increasing the robustness of the estimation when combined with the prior relaxation strategy described in section 3.2.2.

Thus we can interpret $f(\mathbf{y} \mid \mathbf{\Phi}_y)$ as a mixture of multivariate normal distributions, thus $f(\mathbf{y}_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y) = G_{\boldsymbol{\Sigma}_k}(\mathbf{y}_i - \boldsymbol{\mu}_k - \boldsymbol{c}^T \phi(x_i))$, where $G_{\boldsymbol{\Sigma}_k}(x)$ denotes a multivariate zero-mean normal distribution with standard deviation $\boldsymbol{\Sigma}_k$, defined as $G_{\boldsymbol{\Sigma}_k}(x) = ((2\pi)^{\frac{m}{2}} \mid \boldsymbol{\Sigma}_k \mid^{\frac{1}{2}})^{-1} e^{-\frac{1}{2} x^T (\boldsymbol{\Sigma}_k)^{-1} x}$. Hence the *Maximum a Posteriori* (MAP) parameters for $\mathbf{\Phi}_y$ can be found by maximisation of Equation 3.25 giving the following update equations for the model parameters:

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{\boldsymbol{\mu}_k^* \boldsymbol{\Sigma}_k^{*^{-1}} + \left(\sum_{i=1}^n p_{ik} \bar{\mathbf{y}}_i\right) \boldsymbol{\Sigma}_k^{(t)^{-1}}}{\boldsymbol{\Sigma}_k^{*^{-1}} + \left(\sum_{i=1}^n p_{ik}\right) \boldsymbol{\Sigma}_k^{(t)^{-1}}} \tag{3.27}$$

$$\boldsymbol{\Sigma}_k^{(t+1)} = \frac{\sum_{i=1}^n p_{ik} \left(\bar{\mathbf{y}}_i - \boldsymbol{\mu}_k^{(t+1)}\right)^T \left(\bar{\mathbf{y}}_i - \boldsymbol{\mu}_k^{(t+1)}\right)}{\sum_{i=1}^n p_{ik}} \tag{3.28}$$

where the subscript T denotes the transpose, where

$$\bar{\mathbf{y}}_i = \mathbf{y}_i - (\mathbf{c}^{(t)})^T \phi(x_i)$$

and

$$p_{ik}^{(t+1)} = \frac{f(\mathbf{y}_i \mid \mathbf{z}_i = \mathbf{e}_k, \mathbf{\Phi}_y^{(t)}) f\left(\mathbf{z}_i = \mathbf{e}_k\right)}{\sum_{j=1}^K f(\mathbf{y}_i \mid \mathbf{z}_i = \mathbf{e}_j, \mathbf{\Phi}_y^{(t)}) f\left(\mathbf{z}_i = \mathbf{e}_j\right)} \tag{3.29}$$

is the responsibility at the index i and class k, with t denoting the iteration number. The estimation of $\mathbf{c}^{(t+1)}$ under a multimodel scheme is provided in Van Leemput et al. (1999b). To summarise, within the expectation-maximisation framework we include priors on the expected tissue intensities that act to improve robustness in complex segmentation cases.

Spatial Regularisation

Due to the low signal and contrast to noise ratio of the images under analysis, the assumption of statistical independence between neighbouring voxels makes the segmentation prone to noise and image artefacts. A spatial smoothness term by means of a Markov Random Field (MRF) is thus included. This MRF assumes that the probability that voxel i belongs to tissue k depends on its first-order neighbours $\mathcal{N}_i$ and is not dependent on the neonatal image contrast.

Using the mean field approximation as described in Zhang (1992) and Van Leemput et al. (1999b), all previous equations still hold by setting $f(\mathbf{z}_i = \mathbf{e}_k) = f(\mathbf{z}_i = \mathbf{e}_k \mid p_{\mathcal{N}_i}\mathbf{\Phi}_z, \pi_{ik})$ in Equation 3.29, where

$$f(\mathbf{z}_i = \mathbf{e}_k \mid p_{\mathcal{N}_i}, \mathbf{\Phi}_z, \pi_{ik}) = \frac{\pi_{ik}\, e^{-U_{\text{MRF}}(\mathbf{e}_k | p_{\mathcal{N}_i}, \mathbf{\Phi}_z)}}{\sum_{j=1}^{K} \pi_{ij}\, e^{-U_{\text{MRF}}(\mathbf{e}_j | p_{\mathcal{N}_i}, \mathbf{\Phi}_z)}} \tag{3.30}$$

Here, $U_{\text{MRF}}(\mathbf{z}_i \mid p_{\mathcal{N}_i}, \mathbf{\Phi}_z)$ is an energy function dependent on the parameters $\mathbf{\Phi}_z = \{\mathbf{G}, \beta\}$, where $\mathbf{G}$ is an off-diagonal matrix (0 in the diagonal and 1 everywhere else) of elements $\mathbf{G}_{kj}$, measuring transition costs between classes k and j and β is the overall MRF strength, set to 0.5. In this context, π_{ik} normally refers to a spatially variant tissue proportion term, explained in detain in section 3.2.2. Ideally, β should be tweaked depending on the SNR of the image. However, I found the model is relatively insensitive to this value, where a β below 0.1 will result in a very noisy segmentation with large PV areas, and a β above 1 will result in a loss of detail and almost binary probabilistic segmentations. Under anisotropic voxel size, the interaction between neighbours is dependant on a connection strength factor s, introduced as $s = \{s_x, s_y, s_z\} = \{\frac{1}{d_x}, \frac{1}{d_y}, \frac{1}{d_z}\}$, where d is the real-world distance between the centre of neighbouring voxels in each direction. Under this framework,

$$U_{\text{MRF}}(\mathbf{e}_k \mid p_{\mathcal{N}_i}, \mathbf{\Phi}_z) = \beta \sum_{j=1}^{K} \mathbf{G}_{kj} \left(\sum_{l \in \mathcal{N}_i^x} s_x\, p_{lj} + \sum_{l \in \mathcal{N}_i^y} s_y\, p_{lj} + \sum_{l \in \mathcal{N}_i^z} s_z\, p_{lj} \right) \tag{3.31}$$

Relaxation of the Anatomical Priors

Anatomical priors ω that incorporate probabilistic information derived from a digital brain atlas (Kuklisova-Murgasova et al., 2011) are added to the model in order to condition the posterior probabilities and indirectly also condition the model parameters. These atlases are brought into initial correspondence using an affine registration (Ourselin et al., 2000) followed by a free-form non-rigid registration algorithm (Modat et al., 2010). Due to anatomical variability and similarly to Shiee et al. (2011) and Cardoso et al. (2011d), we assume that the proportions π_{ik} are not known *a priori*. Instead, we consider π_{ik} as a sample drawn from a distribution derived from the statistical atlas, i.e., we consider them as a posterior of a Dirichlet distribution. As in Shiee et al. (2011), π_{ik} is updated at each iteration by

$$\pi_{ik}^{(t)} = (1 - \alpha)\omega_{ik} + \alpha(\mathcal{G}_{\sigma_{Der}} * p_{ik}^{(t)}) \tag{3.32}$$

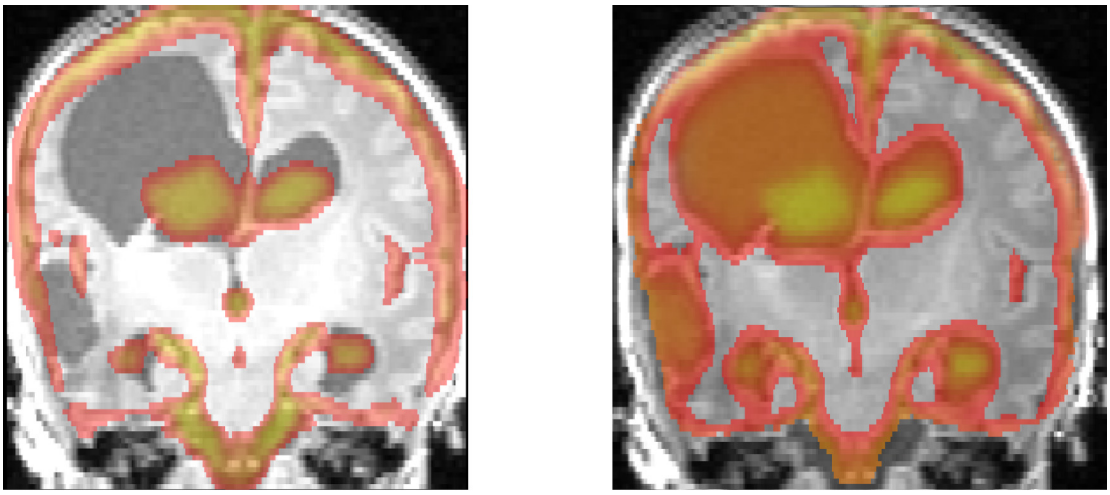


Figure 3.14: The CSF prior a) after non-rigid registration and b) after the iterative relaxation of the priors. Note that the CSF prior now includes the full ventricular area and the right cystic region.

with $\mathcal{G}_{\sigma_{Der}}$ as a Gaussian kernel with standard deviation σ_{Der} and $*$ as the convolution operation, weighted by α. This updating scheme iteratively relaxes π_{ik} spatially, enabling the segmentation of pathological cases. The amount of relaxation is dependent on the parameter α controlling the mixing proportions, and σ_{Der} controlling the amount of regularisation over the posterior. For $\alpha = 0$, only the anatomical priors are used. These anatomical priors ω that incorporate probabilistic information derived from a digital brain atlas are added to the model in order to condition the posterior probabilities and indirectly also condition the model parameters; thus α may be optimised during the segmentation routine and here we adjust the value of α between the initial EM step and the final PV correction step (section 3.2.2) to reflect increased confidence in the segmented data and increase its influence on the PV correction step. Please refer to Figure 3.14 for an example of the anatomical priors before and after relaxation.

Explicit PV modelling

After the adaptive MAP-EM routine, the segmentation is visibly well matched even in cases of marked pathology. However, the procedure is susceptible to producing a layer of WM classified voxels on the GM/CSF, Cerebellum/CSF and brainstem/CSF interfaces (see figure 3.15), a feature specific to neonatal MRI due to the contrast inversion of grey and white matter intensity on both T1-weighted and T2-weighted images and analogous to the equivalent problem in adults (*e.g.* PV on the WM/CSF boundary mapping to GM on T1-weighted images). It is common to assume that if two tissues mix in a voxel, all mixing proportions are equally likely. The PV probability can be seen as a number of mixed Gaussians in between the two pure classes, corresponding to all the possible tissue proportions within a voxel (Van Leemput et al., 2003). Ruan et al. (2000) showed that, for the signal-to-noise ratio levels of current MRI images of the brain, the density of all these PV Gaussian classes can be approximated by a single Gaussian with a small risk ($\alpha < 1$ for D'Agostino-Pearson normality test). Under this assumption we use the previously estimated values of p_{ik}, $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ to initialise the mixed classes so that voxels with

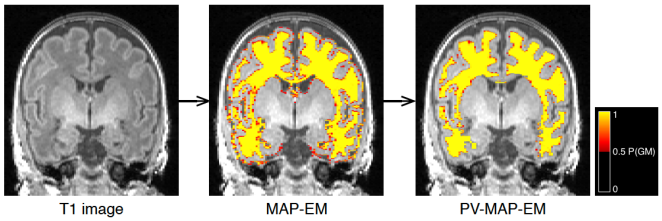


Figure 3.15: From left to right: segmentation pipeline showing the original image, segmentation step with prior alignment and the PV corrected segmentation result. Note the reduced probability of white matter voxels located at the grey matter/CSF boundary after the PV correction step.

ambiguous classification are weighted towards either grey matter or CSF relative to white matter.

Under the assumption of Gaussian distributed classes on log-transformed data, the initial PV class Gaussian parameters can be approximated by a mixed-tissue distribution in each voxel (Kitamoto and Takagi, 1999), with mean equal to the arithmetic weighted average of its composing class parameters weighted by the determinant of the covariance matrix of each class. Thus,

$$\boldsymbol{\mu}^*_{j/k} = \frac{\mid \boldsymbol{\Sigma}_j \mid \boldsymbol{\mu}_j + \mid \boldsymbol{\Sigma}_k \mid \boldsymbol{\mu}_k}{\mid \boldsymbol{\Sigma}_j \mid + \mid \boldsymbol{\Sigma}_k \mid}$$

Due to the multiplicative nature of the probabilities, the mixed class prior is generated as the normalised geometric mean of its composing tissue distributions p_{ij} and p_{ik}.

$$\omega*_{i(j/k)} = \sqrt{p_{ij}\, p_{ik}} \tag{3.33}$$

for mixed voxels and specifically,

$$\omega*_{i(\mathrm{WM})} = \omega_{i(\mathrm{WM})}\left(1 - \sqrt{\omega_{i(\mathrm{GM})}\omega_{i(\mathrm{CSF})}}\right) \tag{3.34}$$

for white matter PV.

This transformation will reduce the *a priori* probability for PV containing voxels to belong to WM. The priors are then normalised in order to sum to one at each voxel position. Even though these new priors cannot be considered as priors in a strict sense as they are derived from patient specific data, they behave as such in the model (Cardoso et al., 2011b). After these patient specific PV priors are created, the image is segmented again using the above described MAP-EM algorithm with the 3 extra PV containing classes (GM/CSF, Cerebellum/CSF and Brainstem/CSF interfaces), resulting in a much reduced amount of PV containing voxels classified as WM (see figure 3.15).

Finally, voxels assigned to the PV classes are converted back to probabilities by calculating the voxel's fractional content (FC) between the two mixing tissue classes, where $FC_{i(j/k)}$ is defined as $\mathrm{FC} = (\boldsymbol{\mu}_j - \bar{\mathbf{y}}_i)/(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)$.

Pipeline

The proposed pipeline can then be summarised in five steps (figure 3.16):

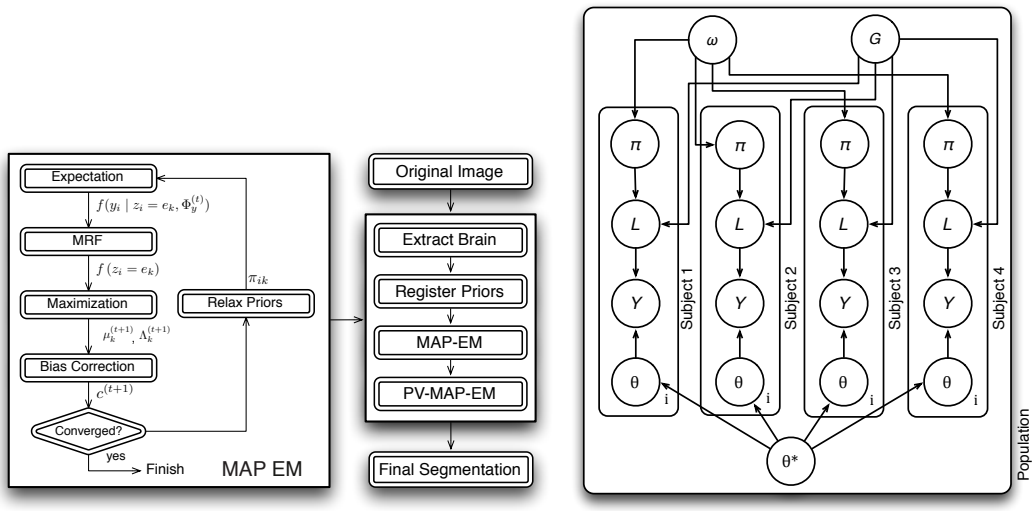


Figure 3.16: Diagram of the proposed adaptive segmentation pipeline and the overall graphical model connecting the different patient specific and population variables.

1. the brain is extracted using the STEPS multi-atlas segmentation propagation scheme (Cardoso et al., 2011c). This method is introduced in section 4.1.

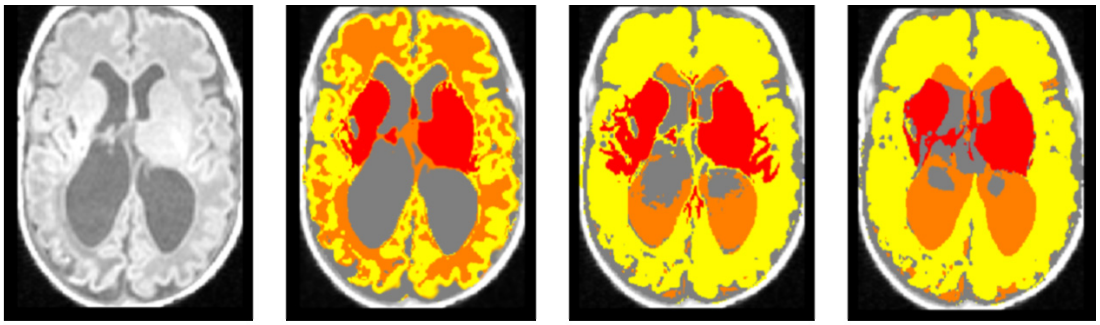
2. registration of the population template and the anatomical priors to the patient space is carried out (Modat et al., 2010).

3. the image used for segmentation is intensity normalised using a robust max-min (assuming 2% outliers) after masking, thus making the image fall in the same intensity range as the population-drawn expected tissue class intensities (Section 3.2.2). The image is segmented into six classes representing the cortical grey matter, unmyelinated white matter, cerebrospinal fluid (CSF) space, cerebellum, deep grey matter (the myelinated white matter is included within this region) and brainstem with the described MAP-EM segmentation algorithm with very broad parameters for the prior relaxation ($\sigma_{Der} = 3$ and $\alpha = 0.5$, enabling a great adaptation of the priors). The estimated bias field is used to correct the intensity inhomogeneity of the image.

4. the bias field corrected image is used to improve the skull-stripping by re-running STEPS.

5. the posterior probabilities from the previous segmentation are used as the priors for the PV-MAP-EM segmentation stage (σ_{Der}=2 and α=0.3 enabling a more prior driven and less adaptive segmentation. Again we segment six classes representing the cortical grey matter, unmyelinated white matter, cerebrospinal fluid (CSF) space, cerebellum, deep grey matter and brainstem.

The chosen values for σ_{Der} and α should intuitively be larger for subjects with divergent anatomical morphologies when compared to a morphologically and pathologically normal subject. In our approach, we chose 2 sets of values for σ_{Der} and α: large values on the first step and smaller ones at the second step. Due to the 2-step procedure, the segmentation becomes much less dependent on the choice of parameters and converges to a sensible solution for both normal and pathological anatomies. The Gaussian



(a) (b) (c) (d)

Figure 3.17: Example (case D in section 3.2.3) of (a) original T1-weighted image and (b) the proposed AdaPT segmentation, followed by (c) the proposed segmentation with prior relaxation but without the priors over the model parameters (section 3.2.2) and (d) the proposed segmentation with priors over the model parameters but without the anatomical prior relaxation (section 3.2.2). The algorithm does not converge to a realistic solution without both improvements. The yellow, red, orange and grey labels represent the cortical and sub-cortical GM, WM and CSF spaces respectively.

parameters are initialised as in Van Leemput et al. (1999b) and section 3.1, by using the priors as the initial estimate of the posteriors (*i.e.* by setting $p_{ik} = \pi_{ik}$) and the bias field coefficients are set to 0.

The proposed pipeline is robust because of both the anatomical prior relaxation, allowing anatomical deviations from an average morphology and the priors over the model parameters which constrain the space of solutions of the algorithm. These two modifications complement each other: the prior relaxation reduces the spatial constraints of the model whilst the priors over the parameters drive the convergence of the model toward population-drawn parameter solutions. An example of the segmentation output alongside examples with each of these improvements applied independently is shown in Figure 3.17. The other model components also contribute to the overall algorithm stability: the MRF introduces spatial smoothness; the bias field correction compensates for INU and the PV correction helps ameliorate misclassification due to the neonate specific GM/WM intensity inversion.

3.2.3 Experiments and results

Data

The data were acquired on a 1.5T Siemens's Avanto. Infants were sedated with an oral dose of chloral hydrate (Rosemont Pharmaceuticals, Leeds, UK) and imaged within a transparent MR-compatible pod. T1-weighted data were acquired with $TR = 17ms$, $TE = 6ms$ and flip angle of 21^o. In total 92 T1-weighted volumes are analysed with resolution of $0.39 \times 0.39 \times 1mm$. The mean gestational age at birth was 27.0 ± 2.7 weeks (range $22.9 - 32.2$ weeks), mean birthweight $966 \pm 380g$ (range $447 - 2470g$) and mean post-menstrual age at scan 40.4 ± 1.74 (range $35.7 - 44.3$). The male to female ratio is $44/48$. The local ethics committee granted permission for this study, and informed parental consent was obtained for each infant. A number of cases have enlarged ventricles (post ischaemia or post haemorrhage) or cystic and diffuse white matter injury and two have transparietal shunts *in situ*. Five cases were selected for manual segmentation on the basis of a variety of brain pathologies associated with prematurity. Case A had a structurally normal brain. Case B had a left parencymal cystic lesion

with ipsilateral widening of the lateral ventricle suggestive of a mature haemorrhagic venous infarction. Case C had a resolved left periventricular haemorrhagic infarction with focal left frontal lobe damage, some cystic damage on the right and consequently a smaller left cerebral hemisphere. Case D had a large right porencephalic cyst a mature haemorrhage within the right posterior limb of the internal capsule and corticospinal tract. The lateral ventricles were enlarged bilaterally with evidence of previous intraventricular haemorrhage, including the fourth ventricle, and loss of white matter bulk. In addition there was mature cystic encephalomalacia within the right temporal, occiptial and parietal lobe with mature haemorrhagic products. Case E had previous intraventricular haemorrhage and a few small periventricular cystic cavities with ventricular enlargement reflecting white matter damage. In addition there was marked increase in extraxial subarachnoid CSF spaces.

Manual segmentation of grey matter

In order to validate the grey matter segmentation, manual segmentation of cortical grey matter was undertaken by a neonatologist with experience of neonatal brain MR imaging. For each case, five blocks of grey matter were manually drawn using ITK-SNAP (Yushkevich et al., 2006) (see Figure 3.19) and the subsequent Dice overlap between these slices and the corresponding slices of the automatic method were obtained. The Dice coefficient, d, is shown in Equation 3.35 where $\mathbf{a}$ and $\mathbf{b}$ are vector representations of the two binary segmentations to be compared.

$$d = \frac{2\sum_i (\min(\mathbf{a}_i, \mathbf{b}_i))}{\sum_i (\mathbf{a}_i + \mathbf{b}_i)} \tag{3.35}$$

The proposed algorithm is compared to a widely-used implementation of Van Leemput et al. (1999b) incorporating an MRF, bias field correction and the 4D anatomical neonatal priors from Kuklisova-Murgasova et al. (2011).

Figure 3.18 presents the overlap of the manual segmentation with automated segmentation for two cases. The block structure of the manual segmentation is clearly visible, thus the Dice score is only calculated over these regions. Relatively normal anatomy in figure 3.18B is visually well-segmented by each algorithm. Estimation of the cortical grey matter in the two pathological cases (see figure 3.18E) is poorer for the standard algorithm than for the adaptive segmentation routine. These results may be compared to figure 3.19 providing a more detailed orthogonal comparison of the manual segmentations in three planes.

Results of the Dice overlap are presented in table 3.2. For the visually straightforward segmentations, the Dice overlap is comparable for each of the suggested algorithm permutations. Advantages of adaptive segmentation become apparent for the pathological cases of ventriculomegaly and extra-axial CSF (figure 3.18E) in which the Dice overlap is substantially higher and the influence of additional priors is higher. For case D, if the initialisation of the spatial priors is too far away, the lack of regularisation in the spatial-prior relaxation approach results in a lower Dice overlap than the ML-EM algorithm. Inclusion of both priors allows the algorithm to converge to a better solution comparable to the accuracy of the more straightforward subjects. For case E the spatial priors are more easily aligned to the morphological

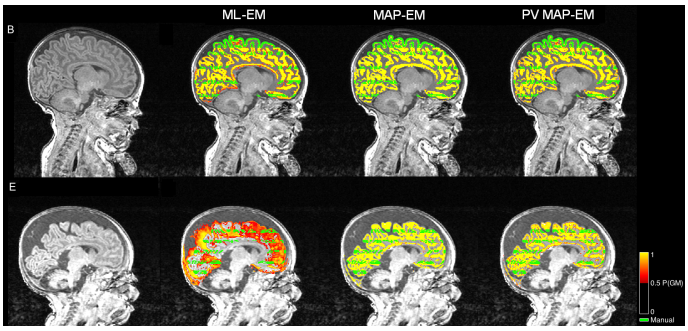


Figure 3.18: Algorithmic differences in segmentation for columns (cases B and E): automatic segmentation using an unmodified ML-EM algorithm; automatic segmentation using the adaptive EM algorithm and automatic adaptive segmentation incorporating a final PV correction step. Manual grey matter segmentation also overlaid in green.

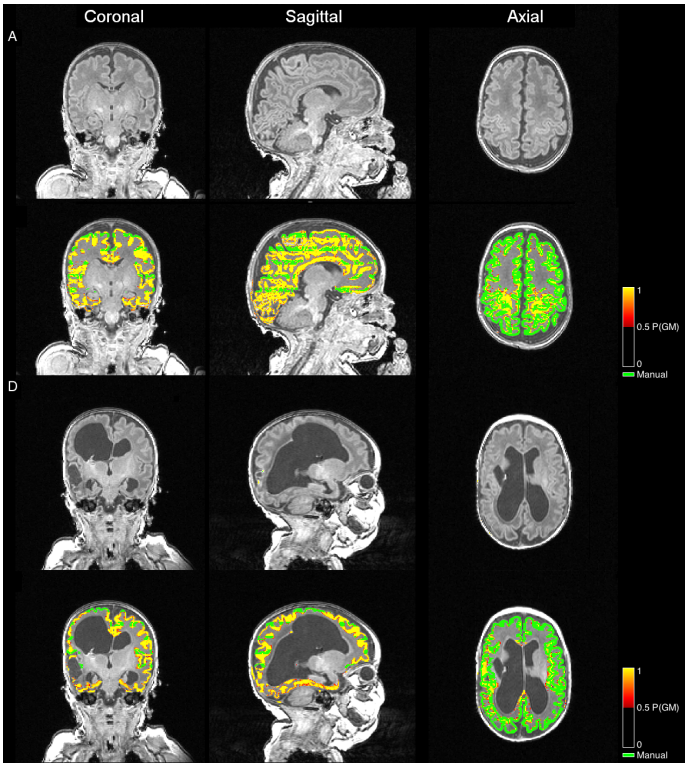


Figure 3.19: Comparison of manual grey matter segmentation with the PV corrected, adaptive EM algorithm for cases A and D.

appearance, thus the atlas relaxation strategy results in a higher Dice score. Combined, these results suggest that the spatial-prior relaxation is advantageous for adapting the priors to morphologically diverse subjects, but without the regularisation introduced by the semi-conjugate prior, the atlas relaxation strategy can converge to poorer solutions than the standard ML-EM.

Manual segmentation of the cerebellum

In order to validate the segmentation of the cerebellum, manual segmentation of the cerebellum for five cases was also carried out. The entire cerebellar volume was manually traced using ITK-SNAP for each case and the subsequent Dice overlap between these slices and the corresponding slices of the automatic

Subject	ML-EM	Adaptive EM (spatial prior relaxation)	EM (intensity prior only)	Adaptive EM (with PV)
A	0.73	0.77	0.78	0.79
B	0.74	0.75	0.76	0.75
C	0.66	0.74	0.75	0.76
D	0.61	0.55	0.60	0.77
E	0.30	0.71	0.63	0.73

Table 3.2: Dice overlap results for comparison with manual grey matter segmentation. Columns correspond to: subject; automatic segmentation using an unmodified ML-EM algorithm (Xue et al., 2007); automatic segmentation using the EM algorithm with spatial prior relaxation only (Shiee et al., 2011); automatic segmentation using EM algorithm with priors on the tissue intensities and lastly, automatic adaptive segmentation incorporating a final PV correction step (AdaPT - figure 3.16).

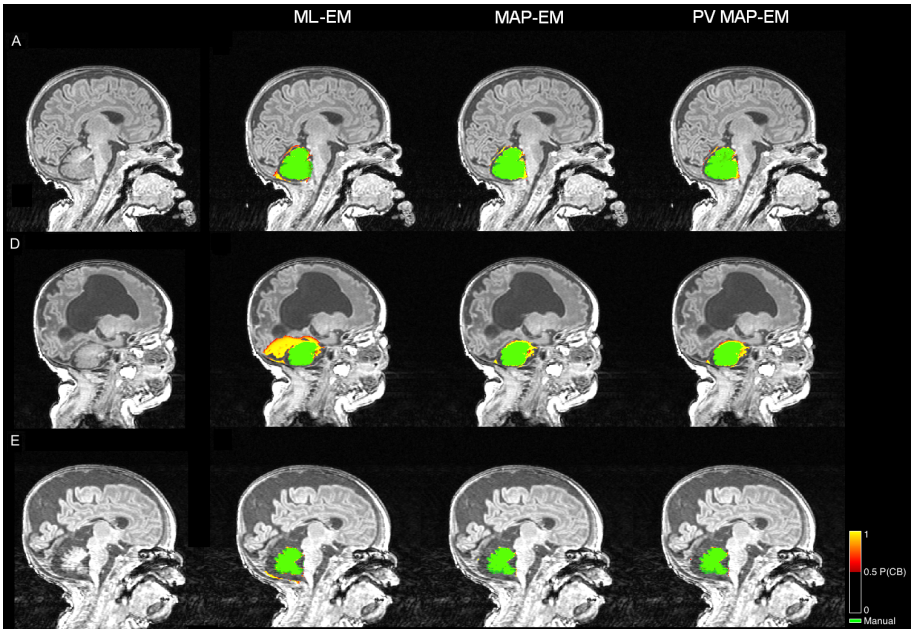


Figure 3.20: Algorithmic differences in segmentation for columns: automatic segmentation using an unmodified ML-EM algorithm; automatic segmentation using the adaptive EM algorithm and automatic adaptive segmentation incorporating a final PV correction step. Manual cerebellum segmentation also overlaid in green.

method were obtained. Cerebellum volume is visually highly variable across the five cases. Figure 3.20 presents the overlap of the manual segmentation with automated segmentation for three cases. Relatively normal anatomy in figure 3.20A is again visually well-segmented by each algorithm although in this case the non-adaptive algorithm does not produce an accurate segmentation of the cerebellum for case D. Estimation of the cerebellum in the two pathological cases is variable: figure 3.20D shows that the standard algorithm performs poorly in this case in the cerebellar region, most likely as a combination of small cerebellum volume and the impact of anatomical distortion as a result of ventriculomegaly. However, standard EM segmentation is able to accurately segment the cerebellum in the case of figure 3.20E due to the excess CSF surrounding the structure.

Results of the Dice overlap are presented in table 3.3, and match the visual summary described for figure 3.20 and the trends shown in Table 3.2 as described for comparison to grey matter manual

Subject	ML-EM	Adaptive EM (spatial prior relaxation)	EM (intensity prior only)	Adaptive EM (with PV)
A	0.82	0.88	0.88	0.90
B	0.87	0.88	0.88	0.89
C	0.69	0.83	0.83	0.84
D	0.48	0.45	0.47	0.83
E	0.79	0.86	0.86	0.88

Table 3.3: Dice overlap results for comparison with manual cerebellum segmentation. Columns correspond to: subject; automatic segmentation using an unmodified ML-EM algorithm (Xue et al., 2007); automatic segmentation using the EM algorithm with spatial prior relaxation only (Shiee et al., 2011); automatic segmentation using EM algorithm with priors on the tissue intensities only and lastly, automatic adaptive segmentation incorporating a final PV correction step (AdaPT - figure 3.16).

segmentation suggesting that the combination of both spatial prior relaxation with a semi-conjugate prior over the image intensities gives optimal performance.

Manual segmentation of the cerebral ventricles

Validation of ventricular volume was also investigated by comparison to manual segmentation of the lateral ventricles. In order to compare just the ventricular region of the automatic CSF class we generate a region of interest from the manual segmentation by dilating it two times using the morphological dilation operation with an adjacent pixel only structure element, thus the Dice score is calculated between the manual and automatic segmentations within this region of interest. This approach is justifiable since the ventricular CSF is always under-segmented rather than over-segmented for the 15 subjects chosen in this validation. Of these 15 cases, 11 have normal anatomy and a further four have marked ventricular enlargement. For each dataset, the Dice score in equation 3.35 was calculated between the proposed method and the manual segmentation.

Figure 3.21 presents the overlap of the manual segmentation with automated segmentation for one normal and one case with ventriculomegaly. Relatively normal anatomy in figure 3.21A is visually well-segmented. Figure 3.21D shows improved ventricle segmentation in the case of ventriculomegaly, which the non-adaptive algorithm fails to accurately segment. For improved clarity, figure 3.21D show the manual ventricle segmentations overlaid on the corresponding non-adaptive, adaptive without PV correction and AdaPT segmentations respectively.

Table 3.4 presents the improvements in Dice overlap when using an adaptive technique. In the cases of reasonably normal anatomy, the improvement in Dice overlap is modest, but none-the-less consistent. However, for cases of ventricular enlargement, the obtained Dice scores are significantly increased when using an adaptive routine relative to the standard EM algorithm. Dice overlap comparison between the two groups is not possible since the Dice overlap is volume dependent, thus higher values should be expected in the case of ventriculomegaly.

Comparison with manual segmentation of unmyelinated white matter

The automatic unmyelinated white matter segmentation is compared to manual segmentation of the unmyelinated white matter on the same slices used for the grey matter validation. The subsequent Dice

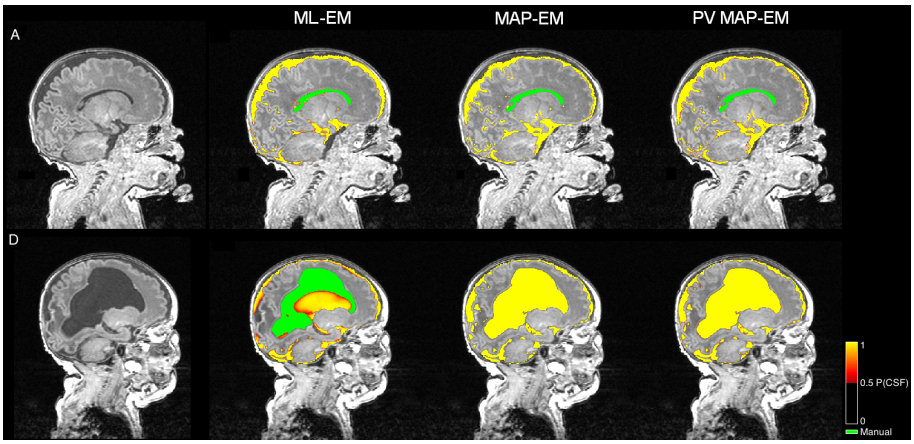


Figure 3.21: Algorithmic differences in segmentation for columns: automatic segmentation using an unmodified ML-EM algorithm; automatic segmentation using the adaptive EM algorithm and automatic adaptive segmentation incorporating a final PV correction step. Manual ventricle segmentation shown overlaid in green for normal (case A) and underlaid for ventriculomegaly (case D).

Subject	ML-EM	Adaptive EM with PV correction
μ(pathological)	0.44±0.10	0.99±0.004
μ(normal)	0.88±0.06	0.94±0.02

Table 3.4: Dice overlap results for comparison with manual ventricle segmentation. Columns correspond to: automatic segmentation using an unmodified ML-EM algorithm and automatic adaptive segmentation incorporating a final PV correction step. Direct comparison between the two groups is not possible since the Dice overlap is volume dependent, thus higher values should be expected in the cases of ventriculomegaly.

overlap between these slices and the corresponding slices of the automatic method are shown in table 3.5. Figure 3.22 illustrates some differences obtained for each method for case C and case D. The low Dice score for the prior relaxation strategy only (Table 3.5 case D, column 2) is the result of using adaptive tissue spatial priors without the intensity prior; this result is also shown in Figure 3.17.

Comparison with manual segmentation of deep grey matter and brainstem

We finalise the validation of the six tissue classes with Dice overlap for whole volume manually segmented brainstem and deep grey matter. Table 3.6 shows the results for the segmentation of the brainstem; Dice scores for this class are slightly lower for the AdaPT method due to under segmentation of the lower brainstem for cases A and B (Example for case C shown in Figure 3.23C-BS). Table 3.7 shows the results for the deep grey matter segmentation; this class also includes any myelinated white matter region. Dice scores for this class are relatively invariant due to the lower distortion in the pathological cases (Figure 3.23C-DGM) relative to the white matter and CSF volumes.

3.2.4 Summary of cohort data

Significant correlations with gestational age at birth are obtained for the white matter, grey matter, deep grey matter, cerebellum and brainstem (the linear correlation coefficients for each group are: $r = [0.55, 0.35, 0.47, 0.40, 0.29]$ respectively, each with $p < 0.01$). The ratio of white to grey mat-

Subject	ML-EM	Adaptive EM (spatial prior relaxation)	EM (intensity prior only)	Adaptive EM (Without PV)	Adaptive EM (with PV)
A	0.85	0.87	0.87	0.86	0.92
B	0.84	0.87	0.86	0.86	0.92
C	0.85	0.87	0.88	0.87	0.93
D	0.30	0.01	0.31	0.76	0.89
E	0.56	0.76	0.64	0.83	0.90

Table 3.5: Dice overlap results for comparison with manual white matter segmentation. Columns correspond to: subject; automatic segmentation using an unmodified ML-EM algorithm (Xue et al., 2007); automatic segmentation using the EM algorithm with spatial prior relaxation only (Shiee et al., 2011); automatic segmentation using EM algorithm with priors on the tissue intensities only and for the final two columns, automatic adaptive segmentation without and with the final PV correction step (AdaPT - figure 3.16).

Subject	ML-EM	Adaptive EM (spatial prior relaxation)	EM (intensity prior only)	Adaptive EM (with PV)
A	0.70	0.74	0.76	0.63
B	0.75	0.77	0.78	0.69
C	0.75	0.77	0.79	0.73
D	0.67	0.59	0.67	0.83
E	0.80	0.79	0.83	0.83

Table 3.6: Dice overlap results for comparison with combined manual brainstem segmentation. Columns correspond to: subject; automatic segmentation using an unmodified ML-EM algorithm (Xue et al., 2007); automatic segmentation using the EM algorithm with spatial prior relaxation only (Shiee et al., 2011); automatic segmentation using EM algorithm with priors on the tissue intensities only and lastly, automatic adaptive segmentation incorporating a final PV correction step (AdaPT - figure 3.16).

Subject	ML-EM	Adaptive EM (spatial prior relaxation)	EM (intensity prior only)	Adaptive EM (with PV)
A	0.86	0.88	0.89	0.85
B	0.86	0.85	0.87	0.84
C	0.80	0.84	0.85	0.84
D	0.85	0.84	0.85	0.81
E	0.79	0.82	0.83	0.86

Table 3.7: Dice overlap results for comparison with manual deep grey matter segmentation. This classification includes myelinated white matter. Columns correspond to: subject; automatic segmentation using an unmodified ML-EM algorithm (Xue et al., 2007); automatic segmentation using the EM algorithm with spatial prior relaxation only (Shiee et al., 2011); automatic segmentation using EM algorithm with priors on the tissue intensities only and lastly, automatic adaptive segmentation incorporating a final PV correction step (AdaPT - figure 3.16).

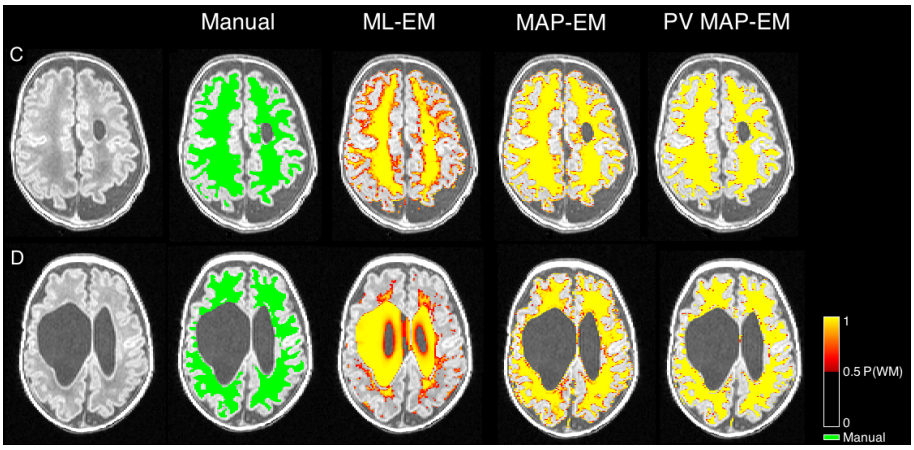


Figure 3.22: Algorithmic differences in segmentation for columns: anatomical reference, manual segmentation, automatic segmentation using an unmodified ML-EM algorithm; automatic segmentation using the adaptive EM algorithm and automatic adaptive segmentation incorporating a final PV correction step. Cases C and D shown.

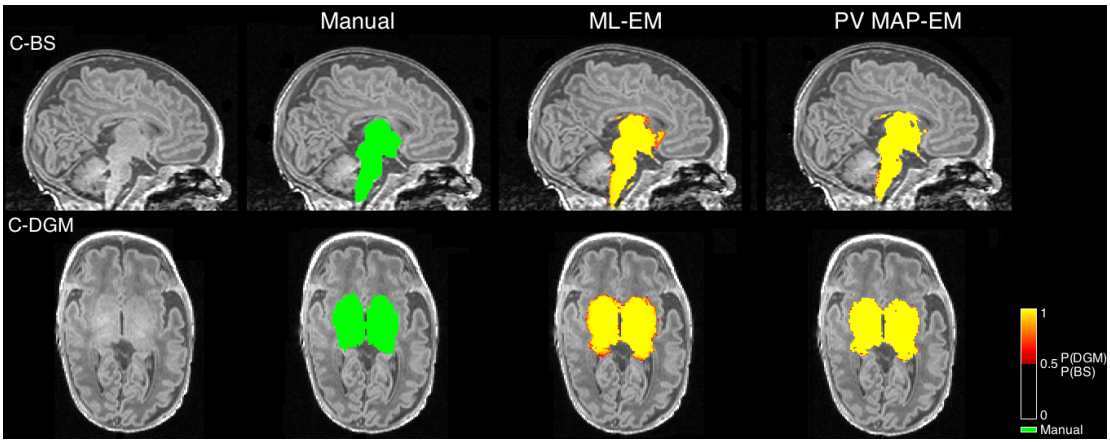


Figure 3.23: Algorithmic differences in segmentation for columns: anatomical reference, manual segmentation, automatic segmentation using an unmodified ML-EM algorithm; automatic segmentation using the adaptive EM algorithm and automatic adaptive segmentation incorporating a final PV correction step. Brainstem and DGM segmentation shown for case C.

ter correlates to gestational age at birth with $r = 0.28$ and $p < 0.01$. Volume of intracranial cerebral spinal fluid does not significantly correlate with gestational age at birth ($p = 0.63$) although CSF as a percentage of brain volume does ($r = -0.34$, $p < 0.01$); increased CSF relative to head size is a common gestational age dependent observation in the preterm phenotype.

Statistics for gestational age at birth in weekly intervals are shown in Figure 3.24 for groupings $(22-23)$ up to $(32-33)$ weeks gestational age at birth with number of cases for each interval: $n = [1, 10, 11, 21, 9, 6, 8, 4, 8, 10, 4]$. Many of the marked outliers correspond to subjects previously discussed in the manual segmentation figures: outliers 1-5 have normal appearance and the automated adaptive segmentation gives good visual performance, in all cases the infants have larger than average birthweight for their category; outliers 6-7 are of birth weight $< 600g$ and thus have low intracranial volume; outlier 8 has a mild increase in extra-axial CSF visible in the imaging data; outliers labelled 9 are

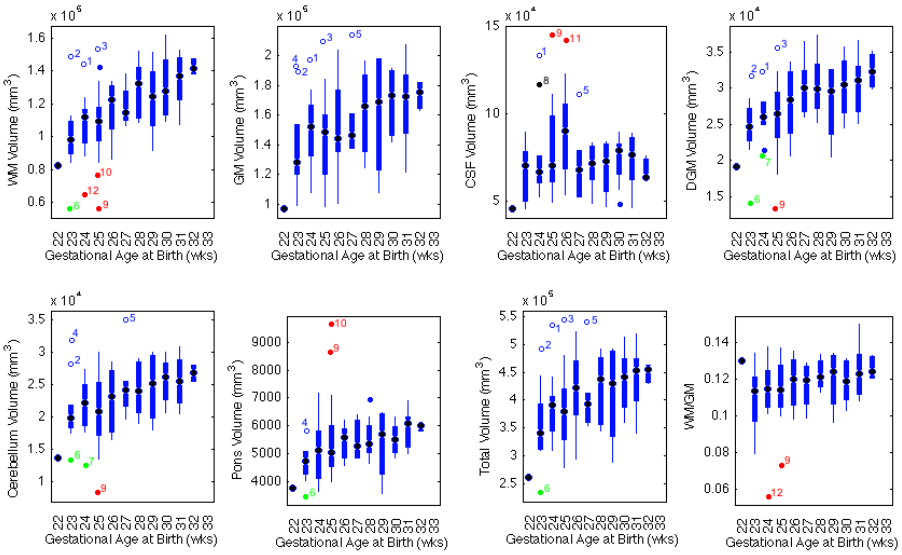


Figure 3.24: Volumetry distribution for automated segmentation at weekly intervals for 22-23 to 32-33 weeks gestational age at birth (see text). Outliers correspond as follows: 1-5 normal appearance and good segmentation quality; 6-7 birth weight $< 600g$; 8 - subject has mildly increased extra-axial CSF; 9 - subject corresponds to subject with bilateral germinal matrix haemorrhage, cortical damage and parenchymal thinning; 10 - subject corresponds to Figure 3.19D; 11 - subject corresponds to Figure 3.18E; 12 - subject corresponds to a subject with a transparietal shunt *in situ* and evidence of previous intraventricular haemorrhage.

values obtained from a subject with bilateral germinal matrix haemorrhage, cortical damage and extensive brain parenchymal thinning (see figure 3.21d); outlier 10 corresponds to Figure 3.19b with ventriculomegaly; outlier 11 corresponds to Figure 3.18b with extra-axial CSF and lastly, outlier 12 corresponds to a subject with a transparietal shunt *in situ* and evidence of previous intraventricular haemorrhage.

3.2.5 Discussion and Conclusion

This work has developed a tool for the accurate segmentation of challenging neonatal MRI with specific emphasis on pathological cases. The method uses a MAP-EM algorithm with a prior relaxation strategy combined with a semi-conjugate prior over the intensities and an explicit PV model in order to mitigate the problem of misclassified PV voxels. Both these additions are important and complement each other; the prior relaxation lessens the spatial constraints whilst the priors over the parameters contribute to maintain the stability of the EM convergence by avoiding local minima. Manual segmentation has provided evidence that the method is more accurate when segmenting challenging cases in comparison to algorithms without these modifications, particularly in the highly variable white matter and CSF regions, but less so in the brainstem and deep grey matter regions. The inclusion of these cases could enable more precise statistics to be inferred relating to patient outcome during infancy and childhood. In particular, the iterative relaxation of the anatomical priors and the priors over the model parameters allows the accurate segmentation of anatomies that are widely divergent, including variation in head-shape, head-size, large variation in CSF volume and pathological morphologies. An accurate and robust segmentation algorithm is also a prerequisite for subsequent processing, for instance when analysing the

cortical surface and thickness or when tracking the progression of myelination. Furthermore, the ability of the algorithm to accurately segment T1 weighted MRI images is highly beneficial for subsequent cortical shape and thickness analysis. It is also important to note that due to the existence of transitive structures, myelination and susceptibility artefacts, the Gaussian distribution used in this work cannot correctly model the underlying tissue at this maturation stage. However, the use of population priors, MRF and semi-conjugate priors over the model parameters in the AdaPT method ensures that the model is constrained and stable enough to cope with distribution outliers.

The manual segmentation used in this evaluation requires some interpretation as a result of a binary classification being imposed on an anisotropic, partial-volumed voxel. The process is quite laborious and thus more extensive manual segmentation on higher numbers of cases or detailed inter (or intra) rater agreement is precluded, thus as with most studies on segmentation algorithms, the interpretation is one of algorithm consistency over accuracy. More detailed inspection of the PV correction scheme might benefit from these repeated manual segmentation strategies, but in this case the work involved outstrips the one or two Dice overlap percentage points gained from these correction methods which result in visibly improved segmentation. Although these extensive analyses are important, the emphasis in this work is the development of a tool for automatic segmentation of marked pathology, and in the examples shown, the most significant advantage is given by the proposed combined strategy.

Of note in the term-equivalent infant brain is the notion of two white matter classes: a dominant unmyelinated white matter class and an advancing myelinated white-matter class. At about term equivalent age, the myelinated class is largely restricted to the projective corticospinal tracts emerging from the deep grey matter structures. With regards to the segmentation, a sub-segmentation of the white matter and deep grey matter classes could be carried out, thus enabling the future investigation of whether the spatial pattern of myelinated and unmyelinated white matter provides a useful biomarker of developmental outcome. A similar methodology could be applied to the automated extraction of cystic regions as a sub-class of the CSF segmentation. Analysis of cystic load or the location of cystic injury might also provide a useful biomarker. These steps are additional post-segmentation steps, although if a suitable algorithm could be found, the information from these additional sub-classes might be incorporated into the iterative adaptive framework. A possible solution might be the use of methods similar to Van Leemput et al. (2001), where myelinated WM, edema and cystics leasion could in theory be considered as model outliers. Furthermore, the accurate segmentation and jointly estimated deformation between the initial priors and the subject may also allow the initial propagation of a finer atlas, for instance of deep grey matter structures, facilitating further analysis of these features across diverse neonatal cohorts.

The adaptive neonatal brain segmentation algorithm described here is implicitly multi-modal, thus allowing T2-weighted data to further improve segmentation accuracy and robustness. This inclusion might also allow the algorithm to accurately distinguish additional white matter or cystic lesion classes pertaining to ongoing myelination or white matter damage in the neonatal brain.

In summary, this section presents a segmentation tool developed specifically for the diverse natural and pathological variability seen in very preterm neonatal brain MRI. Evaluation by manual segmenta-

tion demonstrates the advantages of the adaptive strategy in difficult cases, facilitating the inclusion of these infants in volumetric biomarker studies of novel neuroprotective agents.

Chapter 4

Structural Parcelation

4.1 STEPS: Similarity and Truth Estimation for Propagated Segmentations

Anatomical segmentation of structures of interest is critical to quantitative analysis in medical imaging. Several automated multi-atlas based segmentation propagation methods that utilise manual delineations from multiple templates appear promising. However, high levels of accuracy and reliability are needed for use in diagnosis or in clinical trials. This section introduces a new local ranking strategy for template selection based on the locally normalised cross correlation (LNCC) and an extension to the classical STAPLE algorithm by Warfield et al. (2004), which I'll refer to as STEPS for Similarity and Truth Estimation for Propagated Segmentations. It addresses the well-known problems of local vs. global image matching and the bias introduced in the performance estimation due to structure size.

4.1.1 Derived publications and methodological sumary

- **Cardoso, M. J.** Leung, K. K.; Modat, M.; Cash, D.; Barnes, J.; Fox, N.C.; Ourselin, S.: STEPS: Similarity and Truth Estimation for Propagated Segmentations. (2012) Medical Image Analysis (Under Submission)

- **Cardoso, M. J.**; Modat, M.; Ourselin, S.: Multi-label similarity and truth estimation for propagated segmentations (STEPS) validation (2012) MICCAI MASLF

- **Cardoso, M. J.**; Modat, M.; Cash, D.; Ourselin, S.: Multi-STEPS: Multi-label Similarity and Truth Estimation for Propagated Segmentations. (2012) MMBIA

- **Cardoso, M. J.** Leung, K. K.; Modat, M.; Barnes, J.; Ourselin, S.: Locally Ranked STAPLE for template based segmentation propagation. (2011) MICCAI - MALSF

This section presents a new algorithm that incorporates a local similarity metric to estimate the expected image-based performance of each classifier on a voxel-by-voxel basis into a STAPLE formulation. This is the first time a local ranking and sampling strategy has been introduced into the STAPLE framework. A new Markov Random Field (MRF) model is also introduced and optimised iteratively over the probabilistic labels in order to add spatial consistency and smoothness between the best local classifiers. This

LNCC metric can cope with spatially variant registration accuracy, enabling the use of smaller template databases. Due to the local nature of the algorithm, it is independent of the selected ROI and more robust to INU in MRI images.

To the best of my knowledge, this is the first time a spatially variant similarity term is introduced in a STAPLE framework, enabling the characterisation of both image similarity and human rater performance in a unified manner.

4.1.2 Methods

In this section, I first introduce the mathematical framework and present the original STAPLE algorithm by Warfield et al. (2004). I then introduce the idea of global and local ranking and the subsequent STAPLE model changes. Finally, the full framework is extended to a multi-label scenario.

The STAPLE algorithm

Let an image with N voxels be denoted by $\mathbf{y}$, with each voxel index denoted by y_i. Also, let $\mathbf{t}$ be an indicator vector of size N, again indexed by t_i, representing the hidden binary true segmentation of the object. The value of t_i will be equal to 1 when the structure is present in position i and equal to 0 when the structure is absent in position i. Let the $\mathbf{d}$ be a vector of size R, with each one of its elements $\mathbf{d}_r$ having size N and representing a candidate segmentation of the object of interest obtained either by manual segmentation or an automatic algorithm. This vector $\mathbf{d}_r$ has the same form as $\mathbf{t}$, with 1 and 0 representing the presence and absence of the structure at each position i. In order to parameterise the sensitivity and specificity of each rater, let $\mathbf{p} = (p_1, p_2, ..., p_R)^T$ and $\mathbf{q} = (q_1, q_2, ..., q_R)^T$ represent the sensitivity and specificity of each one of the R candidate segmentations, indexed by j. Here, $\mathbf{p}$ and $\mathbf{q}$ represent a global measure of agreement and disagreement, respectively, between a candidate segmentation and the consensus. Thus, they do not depend on the image index i. In order to estimate $\mathbf{t}$, one needs to maximise the log likelihood of the complete data of this problem $(\mathbf{d}, \mathbf{t})$ given the set of parameters ($\mathbf{p}$, $\mathbf{q}$). Thus, the cost function being optimised is the logarithm of the complete data likelihood $f(\mathbf{d}, \mathbf{t}|\mathbf{p}, \mathbf{q})$, described as

$$(\hat{\mathbf{p}}, \hat{\mathbf{q}}) = \arg \max_{\mathbf{p},\mathbf{q}} \log(f(\mathbf{d}, \mathbf{t}|\mathbf{p}, \mathbf{q})) \tag{4.1}$$

Using the definition of sensitivity and specificity, $\mathbf{p}$ and $\mathbf{q}$ can be described as the "true positive fraction" and "true negative fraction" . Thus, p_j and q_j can be represented by

$$p_j = Pr(d_{ij} = 1|t_i = 1)$$
$$q_j = Pr(d_{ij} = 0|t_i = 0)$$

This model assumes that the candidate segmentations are independent from each other and thus $p_j \perp p_{j'}$, $q_j \perp q_{j'}$ and $d_{ij} \perp d_{ij'}$, $\forall j \neq j'$. Equation 4.1 can thus be maximised by an Expectation-Maximisation algorithm. The notation $w_i^{(k)}$ is used to represent the expected probability of the true segmentation at

voxel i being equal to one at iteration k. Here, $w_i^{(k)}$ is then defined as

$$w_i^{(k)} \equiv f\left(t_i = 1 \mid \mathbf{d}_i, \mathbf{p}^{(k)}, \mathbf{q}^{(k)}\right) \tag{4.2}$$

$$= \frac{a_i^{(k)}}{a_i^{(k)} + b_i^{(k)}} \tag{4.3}$$

with

$$a_i^{(k)} \equiv f(t_i = 1) \prod_j f\left(d_{ij} | t_i = 1, p_j^{(k)}, q_j^{(k)}\right) \tag{4.4}$$

$$b_i^{(k)} \equiv f(t_i = 0) \prod_j f\left(d_{ij} | t_i = 0, p_j^{(k)}, q_j^{(k)}\right) \tag{4.5}$$

and the parameters (p,q) at iteration $(k+1)$ are optimised by

$$p_j^{(k+1)} = \frac{\sum_i w_i^{(k)} d_{ij}}{\sum_i w_i^{(k)}} \tag{4.6}$$

$$q_j^{(k+1)} = \frac{\sum_i w_i^{(k)} (1 - d_{ij})}{\sum_i w_i^{(k)}} \tag{4.7}$$

A more detailed explanation of the model solution can be found in Warfield et al. (2004).

Iterative MRF regularization

Similarly to the original STAPLE algorithm, a Markov Random Field (MRF) is used to add spatial consistency. The MRF model presented in the original STAPLE paper is a post processing step that works on integer labels and not on the probabilities. In order to introduce the MRF spatial consistency within the same optimisation framework, the model presented in Cardoso et al. (2011b) and in section 3.1 is used. This model is not only computationally more efficient that the one presented in the original STAPLE algorithm (Warfield et al., 2004), as it is updated with a mean field approximation, but it works on probabilistic labels and not on the final binarised labels.

This MRF model can be described as a non-binary multi-class extension of the Potts model with the neighbouring clique strength dependent of the voxel size. It has the form

$$f(t_i = k) = \frac{\pi \, e^{-\beta_i U_{\text{MRF}(t_i = k)}}}{\sum_j \pi_j \, e^{-\beta_i U_{\text{MRF}(t_i = j)}}}$$

with

$$U_{\text{MRF}(t_i = k)} = \sum_{j=1}^{R} \mathbf{h}_{kj} \left(\sum_{l \in \mathcal{N}_i^x} s_x \, w_{lj} + \sum_{l \in \mathcal{N}_i^y} s_y \, w_{lj} + \sum_{l \in \mathcal{N}_i^z} s_z \, w_{lj} \right)$$

where $\mathbf{H}$ is a K x K matrix with element h_{kj} containing the transition energy between the class k and the class j, and with the MRF neighbourhood system defined as $\mathcal{N}_i = \left\{\mathcal{N}_i^x, \mathcal{N}_i^y, \mathcal{N}_i^z\right\}$. Here, $\mathcal{N}_i^x$, $\mathcal{N}_i^y$, $\mathcal{N}_i^z$ represent the two direct neighbours of i in the x, y and z directions respectively. Also, π_k is the proportion of the object k in the full image, estimated from w at each iteration and s_x, s_y and s_z are the inverse of the voxel size in the x, y and z directions respectively. As the presented formulation only

has two classes, the MRF matrix $\mathbf{H}$ is set up with the diagonal elements equal to 0 and the off-diagonal elements equal to 1.

When applying this MRF model in a multi-label fusion scenario, the MRF energy function can be extended to incorporate anatomically derived information about the expected neighbourhood transitions as in section 3.1 and in Cardoso et al. (2011b). Conversely, the classical MRF presented by Warfield et al. (2004) assumes that the transition between every pair of classes has the same probability. For the rest of the thesis, β_i is considered constant throughout the image and equal to 0.5. Both the value of β_i and the matrix H can be optimised in order to improve the overall results. Nonetheless, I'll refrain from this optimisation due to it's computational complexity.

Global and Region-of-interest based ranking

In the original STAPLE paper, Warfield et al. (2004) states that implicit in this model is the notion that the experts have been trained to interpret the images in a similar way. The segmentation decisions may differ due to random or systematic rater differences, and a probabilistic estimate of the true segmentation can be formulated as an optimal combination of the observed decisions and a prior model. Thus, these implicit assumptions may not hold when STAPLE is used for segmentation propagation. For segmentation propagation purposes, the errors can come from different morphological characteristics between the images, bad registration results and even the resampling method.

In order to ameliorate this problem, Aljabar et al. (2009) proposed the use of a global normalised cross correlation (GNCC) based metric to rank the registered templates according to the image being segmented in order to only include propagated segmentations that are consistently accurate. Leung et al. (2010) then introduced the same concept in a STAPLE framework, where the GNCC was calculated on a region of interest defined by the union of the propagated labels, resulting in an improved segmentation accuracy. This metric was used because it was shown to provide a good criterion for template selection in multi-centre imaging data (Aljabar et al., 2009). Once a rank of best to worst matches for each template was established, a subset of the highest ranked matchers was used to propagate the template labels onto the images to be segmented. This methodology still has some limitations, because the morphology of the structure and the quality of the registration is characterised as a single global image metric based on the NCC. Thus, in order to provide a good segmentation, either the registration algorithm must perform well in most cases, or the database has to have enough samples with the relevant type of morphology for the image being segmented. For example, if one wants to segment the temporal cortex of a patient's brain using segmentation propagation, the database would have to be large enough to contain enough templates with the same morphological features (*e.g.* number of sulci and giri) as the image to be segmented, so that the registration algorithm can match these features. Also, the registration might work very well in some areas but less well in other areas, leading to an ambiguous NCC value and to the introduction of errors in the label fusion process.

Local ranking for segmentation propagation

Without loss of generality, in this work, the local image similarity between images is assessed using the fast locally normalised correlation coefficient (LNCC), as proposed by Cachier et al. (2003). This is

contrary to what was suggested by Artaechevarria et al. (2009), as I've found better performance with a LNCC based image similarity than with the local mean squared difference (LMSD). Nonetheless, the framework is general enough to allow any metric to be used.

The fast LNCC metric used in this work is similar to the LNCC metric also presented by Artaechevarria et al. (2009) but the mean and standard deviation are calculated on a local Gaussian window using a convolution method. This makes the LNCC estimate smoother and computationally less expensive. Let $\mathbf{y_r}$ represent a propagated intensity image from the atlas after registration and $\mathbf{y}$ represent the target image to be segmented. Under this formulation, the LNCC at position i will be given by

$$LNCC_i = \frac{\langle \mathbf{y}, \mathbf{y_r} \rangle_i}{\sigma_i(\mathbf{y})\sigma_i(\mathbf{y_r})}$$

where

$$\langle \mathbf{y}, \mathbf{y_r} \rangle_i = \mu(\mathbf{y} \cdot \mathbf{y_r})_i - \mu(\mathbf{y})_i \cdot \mu(\mathbf{y_r})_i \qquad \mu(\mathbf{y} \cdot \mathbf{y_r})_i = \mathcal{G}_\sigma * (\mathbf{y} \cdot \mathbf{y_r})$$
$$\mu(\mathbf{y})_i = \mathcal{G}_\sigma * \mathbf{y} \qquad \mu(\mathbf{y_r})_i = \mathcal{G}_\sigma * \mathbf{y_r}$$
$$\sigma_i(\mathbf{y}) = \sqrt{\mu(\mathbf{y}^2)_i - \mu(\mathbf{y})_i^2} \qquad \sigma_i(\mathbf{y_r}) = \sqrt{\mu(\mathbf{y_r}^2)_i - \mu(\mathbf{y_r})_i^2}$$

with $\mathbf{y}^2$ representing a element-by-element multiplication of $\mathbf{y}$ with itself, $*$ denoting the convolution operator, the $\cdot$ here denotes an element by element multiplication and $\mathcal{G}_\sigma$ being a Gaussian kernel with standard deviation σ.

Due to the local nature and smoothness of the metric, the similarity between the images is described on a smooth voxel by voxel basis, enabling a voxel by voxel ranking with reduced discontinuity effect. If, for example, one starts from a set of 15 template images registered to the image under study, one can then calculate how much each one of the template images correlate locally with the image under study and then take only the top 5 templates on a voxel by voxel basis. There are three main advantages to the proposed method compared to using GNCC: first, the global nature of the GNCC metric may be an unrealistic assumption regarding the complex morphology and shape of the object under study. Secondly, it removes the need to create a ROI for GNCC calculation as the metric itself is local. Finally, this method is more robust to the presence of INU in MRI images, as the local nature of the LNCC method obviates the metric bias due to INU.

STAPLE with local ranking

In order to introduce this local ranking information in the previously described STAPLE algorithm, let a new model variable l_{ij} represent an observed cluster assignment that characterises the image similarity. For the sake of simplicity, l_{ij} will be equal to 1 if the image $\mathbf{g}_k$ is in the top X ranked images at position i and equal to 0 otherwise, here X controls the number of images to use locally according to the LNCC.

This new observation l_{ij} can be integrated into the STAPLE framework by altering the model to

$$(\hat{p}, \hat{q}) = \arg \max_{\mathbf{p},\mathbf{q}} \log(f(\mathbf{d}, \mathbf{t}, \mathbf{l}|\mathbf{p}, \mathbf{q})) \tag{4.8}$$

Using Jensen's inequality, the lower bound on the model will be given by

$$Q(\hat{p},\hat{q}) = \sum_T \sum_i f(t_i|\mathbf{d}_i,\mathbf{l}_i,\mathbf{p},\mathbf{q}) log\left[f(\mathbf{d}_i,\mathbf{l}_i|\mathbf{p},\mathbf{q}) f(t_i)\right] \tag{4.9}$$

where $f(\mathbf{d}_i,\mathbf{l}_i|\mathbf{p},\mathbf{q})$ is defined as

$$\begin{aligned} f(\mathbf{d}_i,\mathbf{l}_i|\mathbf{p},\mathbf{q}) = & \prod_j \left[f(d_{ij}|t_i,p_j^{(k)},q_j^{(k)})^{l_{ij}}\right]^{t_i} \\ & \cdot \prod_j \left[f(d_{ij}|t_i,p_j^{(k)},q_j^{(k)})^{l_{ij}}\right]^{(1-t_i)} \end{aligned} \tag{4.10}$$

In this new framework, Eq. 4.4 and 4.5 will now be equal to

$$\begin{aligned} a_i^{(k-1)} &\equiv f(t_i=1) \prod_j \left[f(d_{ij}|t_i,p_j^{(k)},q_j^{(k)})^{l_{ij}}\right]^{t_i} \\ &= f(t_i=1) \prod_{j:d_{ij}=1} \left(p_j^{(k)}\right)^{l_{ij}} \prod_{j:d_{ij}=0} \left(1-p_j^{(k)}\right)^{l_{ij}} \\ b_i^{(k-1)} &\equiv f(t_i=0) \prod_j \left[f(d_{ij}|t_i,p_j^{(k)},q_j^{(k)})^{l_{ij}}\right]^{(1-t_i)} \\ &= f(t_i=0) \prod_{j:d_{ij}=0} \left(q_j^{(k)}\right)^{l_{ij}} \prod_{j:d_{ij}=1} \left(1-q_j^{(k)}\right)^{l_{ij}} \end{aligned} \tag{4.11}$$

and Eq. 4.6 to

$$p_j^{(k)} = \frac{\sum_i w_i^{(k-1)} d_{ij} l_{ij}}{\sum_i w_i^{(k-1)} l_{ij}} \tag{4.12}$$

$$q_j^{(k)} = \frac{\sum_i w_i^{(k-1)} (1-d_{ij}) l_{ij}}{\sum_i w_i^{(k-1)} l_{ij}} \tag{4.13}$$

where a_i, b_i, p_i and q_i are only influenced by the the locations where $l_{ij}=1$, *i.e.* only on the locations where the template image is locally similar to the image to be segmented, and $w_i^{(k-1)}$ is defined as in Eq. 4.2.

In this modification to the classic STAPLE algorithm, q_j and p_j now represent the sensitivity and specificity only in areas where each classifier is considered an expert by the LNCC ranking strategy. This results in a 2 step performance estimation that decouples the two sources of error: one based on the LNCC image similarity metric observation, characterising the non uniform registration accuracy and shape differences and the other step characterising the specificity and sensitivity of each classifier when compared with the consensus classification.

In this algorithm, I use a LNCC ranking-based binary cluster assignment for the observed variable l_{ij}. This approach is analogous to a sampling scheme, where samples with low local similarity are rejected from the fusion. However, the framework allows non-binary cluster assignments, where different samples can have different *importance* weights.

Performance parameter bias due to structure size

In the original STAPLE formulation, the performance parameters are estimated using all the samples from the image. In this formulation, if the size of the object and the size of the background are very different, the algorithms convergence results in both mathematical precision issues (due to the limited floating-point accuracy representation of $\mathbf{q}$ and $\mathbf{p}$) and biased performance parameters. For example, in a situation where the size of the object is much smaller than the background, the specificity q_j will tend to 1 because $\sum_i w_i(1 - d_{ij})$ will be approximately the same as $\sum_i w_i$ as most pixels in the image are $\mathbf{d}_i = 0$. Equally, due to the small size of the object, $\sum_i w_i d_{ij}$ will be much less similar to $\sum_i w_i$ and thus p_j will not be as close to 1 as q_j. This effect can be seen in Warfield et al. (2004) Table I and Table II. When these biased values of p_j and q_j are then used to calculate the new w_{ij}, b_i will tend to 0, and thus w_{ij} will tend to 1. If the STAPLE output w_{ij} is then thresholded at 0.5 confidence, the object will look over-segmented. In order to avoid an over-segmentation effect, one tends to threshold w_{ij} at very high values, *e.g.* a threshold of 0.9999, as used in (Leung et al., 2010). The optimal threshold will depend on many factors like the number of classifiers used, the mean value of $\mathbf{p}$ and $\mathbf{q}$ and even the value of β for the MRF. Also, because the value of wik will be very close but different from 1, numerical precision becomes an issue. Due to all these issues and given that this threshold is normally set to a constant value within the same study (even if more classifiers are used), the performance results of the STAPLE classifier fusion have a characteristic bumpy shape (Leung et al., 2010). Furthermore, the performance peak in terms of segmentation accuracy will depend on the chosen threshold, making all the analysis biased towards this choice.

Rohlfing et al. (2004b) suggested that only updating and using disputed samples for parameter estimation can improve the computation time. One should note that this approach not only improves the computation time, but most importantly, it also improves segmentation performance by reducing the $\mathbf{p}$ and $\mathbf{q}$ unbalance and consequently the numerical precision issues. Thus, instead of trying to empirically set a threshold on the STAPLE probabilistic output, I restrict the parameter optimisation to non-consensus voxels. Thus, all the voxels where d_{ij} is equal to either 0 or 1 for all experts j are removed from the estimation. This method assumes that if all the classifiers agree on a label at a certain spatial position i, then the voxel is marked as solved and is not taken into account for the estimation of p_j and q_j. In this case, p_j and q_j represent the sensitivity and specificity only in ambiguous voxels, thus ameliorating the bias caused by structure size. One can then threshold w_{ij} at 0.5 without causing over-segmentation of the object. The effect of this step in terms of the shape of the performance results curve will be shown in the validation section.

In summary, the proposed method, named STEPS (Similarity and Truth Estimation for Propagated Segmentations), can be described as a combination of the LNCC ranking, the MRF and two STAPLE modifications regarding both the introduction of the local indicator function l_{ij} and the removal of consensus voxels from the parameter estimation.

Multi-label extention

Let $\mathbf{t}$ be an indicator vector of size N, indexed by t_i, representing the hidden true label describing several objects under analysis. This hidden label is denoted by an integer value $\{1, 2, \ldots, c\}$, with each value representing a different object of interest, from a total of c objects. Now, let $\mathbf{d}$ be a vector of size R, with each one of its elements $\mathbf{d}_r$ representing a candidate segmentation of the object of interest obtained either by manual segmentation or an automatic algorithm.

In order to extend the concept of sensitivity and specificity of a segmentor j into a multi-class model, a confusion matrix $\mathbf{N}_j$ and its row normalised equivalent λ_j, similar to the ones presented in Xu et al. (1992) and Rohlfing et al. (2004b), are introduced in the notation. The matrix $\mathbf{N}_j$ is defined as

$$\mathbf{N}_j = \begin{pmatrix} n_{11} & n_{12} & \ldots & n_{1c} \\ n_{21} & n_{22} & \ldots & n_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ n_{c1} & n_{c2} & \ldots & n_{cc} \end{pmatrix}$$

with each element $n_{a,b}$ denoting that n samples of class a have been assigned a label b by segmentor j. The elements of the matrix λ_j are then defined as

$$\lambda_j(a, b) = \frac{N_j(a, b)}{\sum_c N_j(a, c)}$$

Similarly to Rohlfing et al. (2004b), using the new definition of the performance parameter λ_j, the posterior probability for sample $\ddot{\imath}\ddot{\imath}_{\dot{\iota}\frac{1}{4}}$ to belong to class c will then be

$$w_{ia} = \frac{f(t_i = a) \prod_j \lambda_j(a, d_{ij})^{l_{ij}}}{\sum_c f(t_i = c) \prod_j \lambda_j(c, d_{ij})^{l_{ij}}}$$

and the performance parameter matrices λ_j are updated at each iteration by setting

$$\lambda_j(a, b) = \frac{\sum_{i:d_{ij}=b} w_{ia} l_{ij}}{\sum_i w_{ia}}$$

In a multi-label scenario, instead of thresholding the output of $\mathbf{w}$ at a certain value, the label with the highest value of w_{ia} at each position i is considered the optimal label.

4.1.3 Validation

The validation of the proposed method is divided into two components, one for single label fusion and one for multi-label fusion. As the availability of manual segmentations is much greater for single labels, the validation of the proposed method for label fusion of individual structures was performed in five steps:

1. The method was applied to synthetic data to show the effect of STEPS on a simulated data set with different morphological properties.

2. STEPS was validated against other label fusion techniques using leave-one-out cross validation.

This validation method was also used to optimise the model parameters.

3. Using the optimised model parameters, a leave-one-group-out validation (jackknifing) was done to demonstrate robustness to simulated database size reduction.

4. Validation was then done on a subset of the publicly available ADNI database in order to show robustness to different atrophy states, different MRI pulse sequences and different scanners.

5. STEPS was finally applied to the ADNI database to show volumetric group differences.

Due to the limited availability of template databases with multi-label s, only one validation step was performed for this scenario. Here, the performance of STEPS was compared to a fusion algorithm specifically tailored to the multi-label scenario called MAPER (Heckemann et al., 2010) when segmenting a set of 30 brain images with 83 manually segmented structures using a leave-one-out cross validation.

Phantom Validation

In order to validate the advantages of local ranking versus global ranking under a constrained experiment, a set of 6 simulated anatomical images with corresponding ground truth labels was generated. Each image represents a highly folded structure similar to the cortex, with the simulated intensities in line with anatomical T1 weighted MRI images. Rician noise was then added to the simulated anatomical images by adding Gaussian noise to both real and complex components in the Fourier domain. These 6 images have different number of gyri, representing different morphologies of the brain. One of these images was chosen as the image to segment and the other 5 were used as a template database. In order to simulate mis-registrations, 3 small random deformation fields were generated and applied per template (see Fig. 4.1(top right)), resulting in 15 different templates with 5 different morphologies, each one with a corresponding label. The proposed method's segmentation was compared to the GNCC-ranked STAPLE (Leung et al. (2010)) using the Dice score as a performance metric. This metric is defined as in section 3.2. For both the method proposed by Leung *et al.* and STEPS, I took the top 5 templates ranked globally (according to the GNCC) and locally (according to the LNCC) respectively.

Results are shown in Fig. 4.1. Using a leave-on-out cross validation, the mean Dice score for STEPS and the Leung *et al.* based method was 0.939 and 0.753 respectively.

Hippocampal segmentation

This section validates the performance of the proposed technique for hippocampal segmentation. A previously described hippocampal template library of manually segmented regions, from 55 subjects, was used (Barnes et al., 2008). The subjects in the template library included 36 subjects with clinically diagnosed AD and 19 controls who had a mean age of approximately 70 years. All scans were acquired at a single site 1.5 T GE scanner using a volumetric T1-weighted sequence. The left and right hippocampal regions were segmented by an expert segmentor. In order to increase the template database size, each image and its flipped mirror image were used as templates, resulting in 110 templates with associated segmentations.

In order to assess STEPS, I performed a leave-one-out segmentation validation on all the images. For each image, the remaining 109 templates were first affinely registered (12 DOFs) using a block

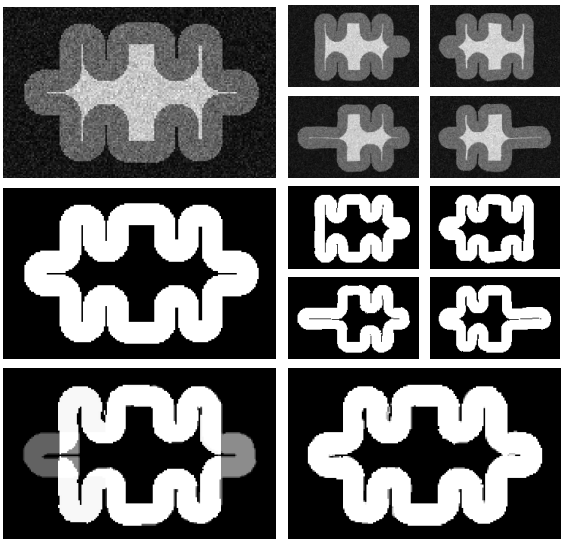


Figure 4.1: From left to right: (Top) The image to segment, four samples from the simulated template database. (Centre) The ground truth segmentation and the respective lables from the template database with different morphologies and simulated registration errors. (Bottom) The probabilistic segmentation using the method from Leung *et al.* (left) and STEPS (right). Note that the lack of local matching has limited the ability of the GNCC method to capture the local features due to the morphologicaly restricted database.

matching approach (Ourselin et al., 2000, 2001) and then non-rigidly aligned using a fast free-form registration algorithm (Modat et al., 2010) to the image under study. The resulting transformations were used to propagate the manual segmentations to the image under study and resampled using nearest-neighbour interpolation in order to maintain their binary nature. The proposed method was compared with the ROI normalised cross correlation (ROINCC) based ranking under a majority voting and STAPLE fusion strategies as proposed by Aljabar et al. (2009) and Leung et al. (2010) respectively, thus assessing the merit of adding the local ranking strategy. In addition, STEPS was also compared with the LNCC weighted voting presented in Yushkevich et al. (2010) and the MSD weighted voting presented in Artaechevarria et al. (2009) in order to give further insights on STEPS performance compared to previously published methodologies. For the sake of completeness, the proposed method was also compared with STEPS without the consensus voxels rejection, as in Cardoso et al. (2011c), in order to assess the improvement due to solving the EM only in areas with uncertainty and also to STEPS without the MRF regularisation, named STEPS-noMRF.

Note that all these comparisons only test the merit of the fusion strategy and not the performance of the full pipeline, as all the templates are registered in the same manner.

Parameter Optimization and Algorithm Comparison

In order to optimise the fusion parameters, the Dice score between the estimated segmentation and the manual segmentation was calculated for different values of Gaussian kernel size, number of labels used and registration parameters. The parameters for the methods proposed by Yushkevich et al. (2010) and Artaechevarria et al. (2009) were not optimised and were defined as suggested by the authors, with $\alpha = 1$ and $\sigma = 1.2mm$ for the former method and $p = -6$ and $r = 5$ for the later method.

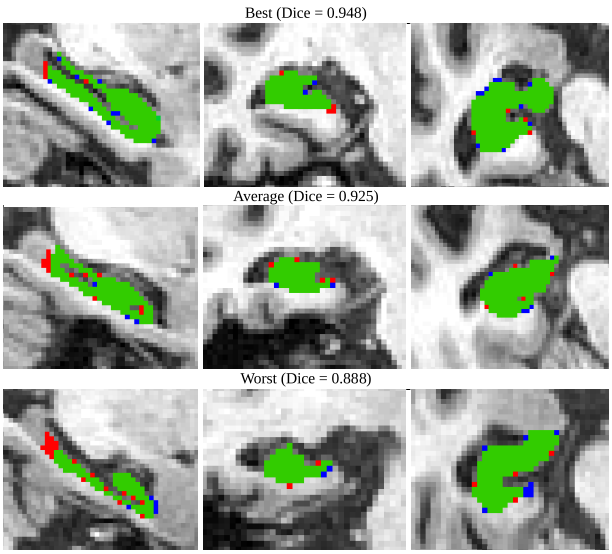


Figure 4.2: Segmentation results showing the best, an average and the worst result. The blue, red and green colours represent the ground truth, the proposed method and the overlap between both segmentations respectively.

For all other methods, I took the top X ranked images, with X varying between 3 and 25. Only odd numbers of X were used in majority voting to avoid bias due to voting ties. For the LNCC ranking the images were locally ranked by setting $l_{ij} = 1$ if the registered template k was in the top X ranked images at position i and to 0 otherwise. For the LNCC ranking in STEPS the value of σ was varied between $1mm$ and $2mm$ with an increment of $0.25mm$ and between $2mm$ and $6mm$ with an increment of $1mm$, for each value of X, in order to find the optimal Gaussian kernel size.

Regarding the other parameters, the region of interest in Aljabar et al. (2009) and Leung et al. (2010) was defined as the union of all the propagated labels dilated D times. The parameter D was also optimised. For each value of X, D was varied between 1 and 4.

The registration parameters were not optimised within the same scheme due to computational complexity. They were only visually optimised on a subset of 10 images in order to produce good registration accuracy. The optimal registration parameters were found to be 2.5mm control-point spacing with 1% bending energy as regularisation.

In total, 324 combinations of parameters, type of ranking and label fusion method were used on each one of the 110 images in the database, totalling 35640 label fusions. Due to the overestimation explained in Section 4.1.2, a constant threshold of 0.9999 was used for all the STAPLE based methods, in order to obtain the final binary segmentation. This threshold is identical to the one used in Leung et al. (2010). For STEPS and weighted voting methods, the threshold was set to 0.5 due to their unbiased nature.

In order to assess the accuracy of the segmentation, the Dice score was calculated between the ground truth manual segmentation and the obtained binary segmentation. The Dice scores for different parameters using STEPS are shown in Fig. 4.3 and an example segmentation from STEPS is shown in Fig. 4.2. The optimal parameters are shown in Fig. 4.4. For STEPS with and without MRF regularisa-

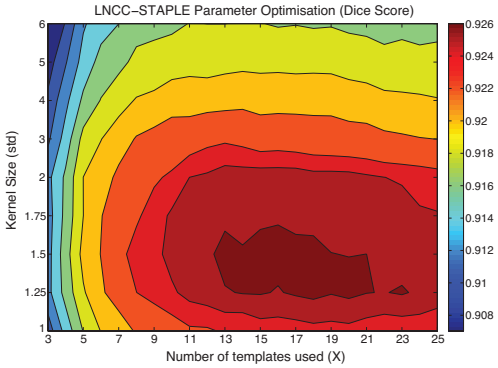


Figure 4.3: Mean Dice score for varying values of σ and X for the proposed STEPS method usign a leave-one-out cross validation. The best parameters were found to be $X = 15$ and $\sigma = 1.5$, with a mean Dice score of 0.925 for STEPS

Table 4.1: Leave-one-out cross validation statistics for different ranking methods and fusion approaches: STEPS, STEPS without MRF (STEPS-noMRF), STEPS with all samples including consessus areas (STEPS-Cons), and the methods in Aljabar et al. (2009), Yushkevich et al. (2010), Leung et al. (2010) and Artaechevarria et al. (2009).

Fusion Method	STEPS		STEPS-noMRF	STEPS-Cons
X,D,σ	15,-,1.5		15,-,1.5	9,-,1.5
Mean	0.925		0.919	0.921
SD	0.014		0.018	0.014
Median	0.929		0.918	0.922
IQ Range	0.018		0.022	0.020
10th Perc.	0.907		0.902	0.904
90th Perc.	0.943		0.937	0.939
t-test p-value	-		$< 10^{-4}$	$< 10^{-4}$
95% CI	-		0.004-0.007	0.004-0.006
Fusion Method	Aljabar	Yushkevich	Leung	Artaechevarria
X,D,σ	6,2,-	-	6,2,-	-
Mean	0.907	0.918	0.909	0.915
SD	0.016	0.014	0.015	0.016
Median	0.909	0.919	0.913	0.913
IQ Range	0.022	0.017	0.018	0.019
10th Perc.	0.886	0.897	0.890	0.893
90th Perc.	0.924	0.933	0.926	0.926
t-test p-value	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
95% CI	0.016-0.020	0.005-0.008	0.009-0.01	0.005-0.01

tion, the optimal parameters are $X = 15$ and $\sigma = 1.5$ (Mean Dice=0.925 and 0.919 respectively). The parameters $X = 9$ and $\sigma = 1.5$ (Mean Dice=0.921) are optimal for STEPS without the consensus voxels rejection as in for the fusion approach, as in Cardoso et al. (2011c). The parameters $X = 6$ and $D = 2$ (Mean Dice=0.909) and $X = 17$ and $D = 2$ (Mean Dice=0.907) are optimal for the fusion approach in Leung et al. (2010) and in Aljabar et al. (2009). Finally, as stated above, the parameters for the fusion approaches in Yushkevich et al. (2010) and Artaechevarria et al. (2009) were not optimised. The mean Dice score was 0.918 and 0.915 respectively. These optimal parameters are used for all comparisons. The Dice score statistics for all methods are shown on Table 4.1 and Figure 4.4.

Using a two tail unequal variance paired t-test, STEPS performed significantly better ($p < 10^{-4}$) than all the other ranking and label fusion strategies for hippocampal segmentation. Confidence intervals for the mean differences, shown in Table 4.1, were found assuming normality of the paired differences.

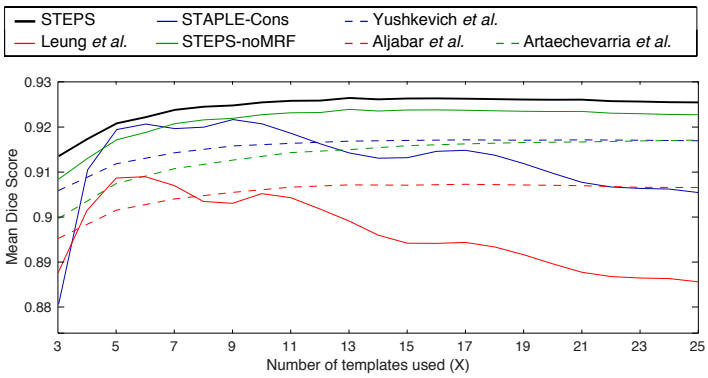


Figure 4.4: The mean Dice score for the full data set for varying values of X with optimal σ and D parameter. The yellow line shows the Dice score when using the naive strategy of fusing all the propagated segmentations with STAPLE. Labels are described in Table 4.1.

Interestingly, the standard deviation of the Dice score did not increase between STEPS and the regionally ranked fusion algorithms. STEPS achieves very high Dice score (0.907) for the 10th percentile data, with the worst segmentation having a Dice score of 0.888. For comparison, the method by Leung et al. (2010) and by Aljabar et al. (2009) only achieved Dice scores of 0.886 and 0.890 respectively for the 10th percentile and a Dice scores of 0.819 and 0.830 respectively for the the worst segmentation. Furthermore, the proposed method (STEPS) has a Dice score equal or higher than all other methods for all data sets. The locally ranked majority voting based methods by Yushkevich et al. (2010) and Artaechevarria et al. (2009) show improved results when compared to both Leung et al. (2010) and Aljabar et al. (2009). However, when compared to STEPS, they still perform significantly worse.

Robustness to database size reduction

One of the main caveats of global ranking methods is the implicit necessity to have a large database in order to be able to represent the population's global anatomical variability. Conversely, STEPS describes image similarity on a local manner. Intuitively, this means that fewer templates are needed to describe the global anatomical variability of a population, as each template contributes locally to the global anatomical variability.

In order to test this hypothesis, I used the same data set as before. However, instead of using a leave-one-out approach, I used a subset of the available template database (110 template) by selecting a smaller set of templates randomly (jackknifing). This is done in order to study the effect of reducing the size of the template database on the results. Assuming a simulated template database of size R, for each data set in the original database, 10 sets of R samples were randomly selected from the remaining 109 templates. Each one of these 10 sets was then considered as a simulated database of size R used to segment the data set under study. The optimised parameters described in Section 4.1.3 were used in order to obtain the fused segmentations. For the sake of comparison, STEPS was compared to the method by Leung et al. (2010) and also to STEPS without excluding the consensus areas (STEPS-Cons).

The degradation was tested at 3 different levels of R (30, 60 and 90), with X varying between 5 and 25 (sampled only at odd values). Thus, 36300 fusions were performed for each method, producing 10

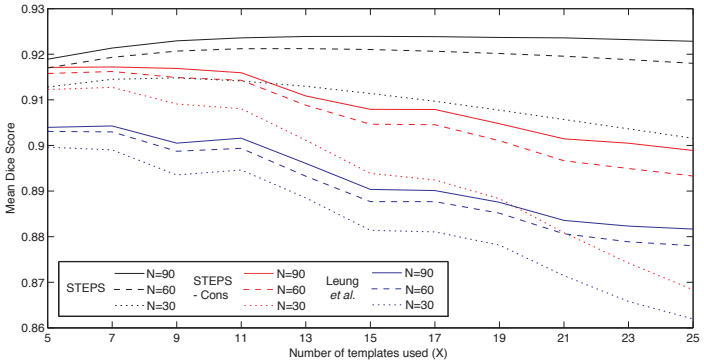


Figure 4.5: The mean Dice score for varying values of X on a simulated database of size R=90, R=60 and R=30 for STEPS, the method by Leung et al. (2010) and STEPS without excluding the consensus areas (STEPS-Cons). Note that there is almost no performance deterioration for the STEPS algorithm between a database of size 90 and 60. STEPS also performs significanly better using only 30 templates than the method proposed by Leung et al. (2010) using the full database.

segmentations per data set, per value of X and per value of R. The resulting Dice score are presented in Fig. 4.5. Using an unequal variance paired t-test to compare the Dice scores, STEPS performed significantly better ($p < 10^{-4}$) using only 30 templates than the ROINCC method using the full database.

Validation on a subset of the ADNI database

In order to characterise the accuracy of using a predefined template database to segment data sets from another database, an expert segmentor manually delineated the left hippocampus on the baseline and repeat T1-weighted MR images of 30 randomly selected subjects. The data consists of 10 Alzheimer's disease (AD), 10 Mild Cognitive Impairment (MCI) and 10 controls, from the ADNI data set. Representative imaging parameters were TR = 2400 ms, TI = 1000 ms, $TE = 3.5ms$, flip $angle = 8^{o}$ with either a $1.25 \times 1.25 \times 1.2mm^3$ or a $0.94 \times 0.94 \times 1.2mm^3$ voxel resolution. The T1-weighted volumetric scans were already pre-processed using the standard ADNI pipeline, including post-acquisition correction of gradient warping, B1 and INU correction and phantom based scaling correction.

Using the previously optimised parameters, 5 different label fusion methods (STEPS, STEPS-Cons, Leung et al. (2010), Yushkevich et al. (2010) and Aljabar et al. (2009)) were used to segment the same data sets. Segmentation accuracy was accessed by calculating the Dice score between the manual and automated segmentations. Results are shown in Table 4.2. Statistical differences were calculated using a two tail unequal variance paired t-test. The mean (SD) Dice score for STEPS was 0.903 (0.021), significantly higher that both Leung et al. (2010) and Aljabar et al. (2009) fusion methods at $p < 10^{-4}$ and higher than the method by Yushkevich et al. (2010) at $p = 0.006$.

Hippocampal measures on the full ADNI data-set

In this section, the ADNI data sets were used to assess both hippocampal volume and change in volume over time (atrophy rate). As suggested in Lötjönen et al. (2011), in order to add PV information to the binary hippocampal segmentation and thus increasing statistical power, each image was also segmented using LoAd (Cardoso et al., 2011b). Hippocampal volume was considered as the sum of the GM fractional content at each voxel position within the binary segmentation obtained from STEPS, multiplied

Table 4.2: Dice score statistics for hippocampal segmentation on 30 ADNI subjects using STEPS, STAPLE-LNCC, STAPLE-ROINCC, MV-LNCC and MV-ROINCC fusion techniques using the previously optimised X (# ranked), D (# dilations) and σ (kernel size in *mm*) parameters.

Fusion Method	STEPS	STEPS-Cons	Leung	Aljabar	Yushkevich
X (# Ranked)	15	7	7	17	19
D (# dilations)	-	-	2	2	-
σ (in *mm*)	1.5	1.5	-	-	1.5
Mean	0.903	0.894	0.869	0.869	0.892
SD	0.019	0.024	0.022	0.021	0.016
Median	0.907	0.901	0.873	0.873	0.895
IQ Range	0.022	0.021	0.028	0.027	0.016
10th Perc.	0.881	0.864	0.844	0.840	0.872
90th Perc.	0.921	0.912	0.894	0.893	0.911

Table 4.3: Subject demographics of the ADNI data set. Mean (SD) unless specified otherwise.

	Controls	MCI	AD
# data sets	200	335	147
Gender, # male	106	213	78
Age, years	76.0 (5.1)	74.9 (7.2)	75.3 (7.3)
Scan.Interval, days	396.3 (46.0)	396.3 (24.3)	390.1 (22.6)
TIV, ml	1584 (144)	1567 (149)	1554 (154)

by the voxel size. The volumes of the left and right hippocampi were added together to give "total" hippocampal volume for each subject.

For the sake of comparison with previously published studies, only the baseline and 12-month repeat volumetric T1-weighted MR scans acquired using 1.5 T scanners were used. In total, 682 subjects were used (200 controls, 335 MCI and 147 AD). The scans were pre-processed following the standard ADNI pipeline, summarised in Leung et al. (2010). Demographics are shown in Table 4.3.

Linear regression was used to assess differences in volumes and change in volumes across groups. The volume, calculated as described above, is considered as dependent observed data. For cross-sectional analysis, the metadata available from the ADNI database comprising of age and gender was used as independent confounding variables. The total intracranial volume (TIV), obtained automatically using SPM8 as described in Leung et al. (2010), was also considered as a confounding variable. For the longitudinal assessment, the atrophy rate was estimated by measuring the difference in volume between baseline and repeat scans normalised by the baseline scan. Because the number of days between baseline and 1-year scans was different between subjects, this information was additionally used as a confounding variable.

The results are shown in Figure 4.6. Statistical differences were calculated using a two tail unequal variance t-test and the significance level was set to $p < 10^{-3}$ due to the intrinsic pathological variability.

The cross-sectional study shows statistically significant hippocampal volumetric differences between the different disease groups. The mean volumes were also similar to previously estimated manual and automatic volumes. For the longitudinal study, even though atrophy rates were not derived directly from the registered serial MR images or propagated from baseline to repeat, the accuracy of the proposed method enables a direct comparison between the volumes of the hippocampus at baseline and 12-month

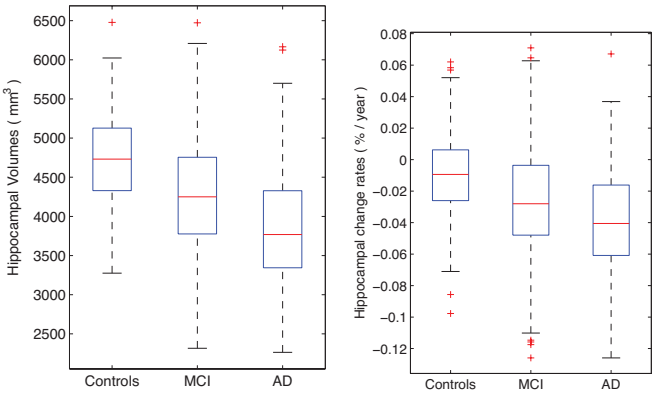


Figure 4.6: Cross-sectional and longitudinal study on 682 data sets from the ADNI database. Left: Total hippocampal volume (left+right side) at baseline; Right: Hippocampal atrophy per year as a percentage of the baseline volume.

Table 4.4: Hippocampal volumes and change rates

		Controls	MCI	AD
Volumes (mm^3)	Mean	5195	4786	4427
	Median	5152	4733	4218
	SD	656	781	903
Change Rates (%/year)	Mean	1.09	2.74	4.04
	Median	0.98	2.61	3.95
	SD	3.0	3.5	3.6

follow up. Results shown in Figure 4.6 and Table 4.4 show statistically significant differences in the mean atrophy rate between disease groups.

Multi-label segmentation propagation and comparison with MAPER

The limited availability of template databases with multi-label s does not allow as complex a validation as with the single label scenario. Thus, only one leave-one-out cross validation was performed, making the validation anecdotal for untested morphologies and severe pathological cases. A previously described template library of 83 manually segmented regions from 30 subjects was used (Hammers et al., 2003, 2007). The median age of all subjects was 31 years, ranging from 20 to 54 years, equal gender distributions and 83% right handed subjects. Scanner parameters are described in Hammers et al. (2007). In order to assess the accuracy for brain using STEPS, I performed a leave-one-out segmentation validation on all the datasets. Each image was first skull stripped using the method proposed by Segonne et al. (2004). Then, for each one of the 30 datasets, the remaining 29 templates were first affinely registered (12 DOFs) using a block matching approach (Ourselin et al., 2000) and then non-rigidly aligned using a fast free-form registration algorithm (Modat et al., 2010) to the image under study. The manual segmentations were then propagated using the previously estimated transformations and resampled using nearest-neighbour interpolation in order to maintain their binary nature. I compare STEPS to a previously published state-of-the-art method called MAPER (Heckemann et al., 2010). STEPS was also tested without the MRF in order to show the improvements in accuracy and smoothness. In order to provide a fair comparison between methodologies, the results presented in the original MAPER paper

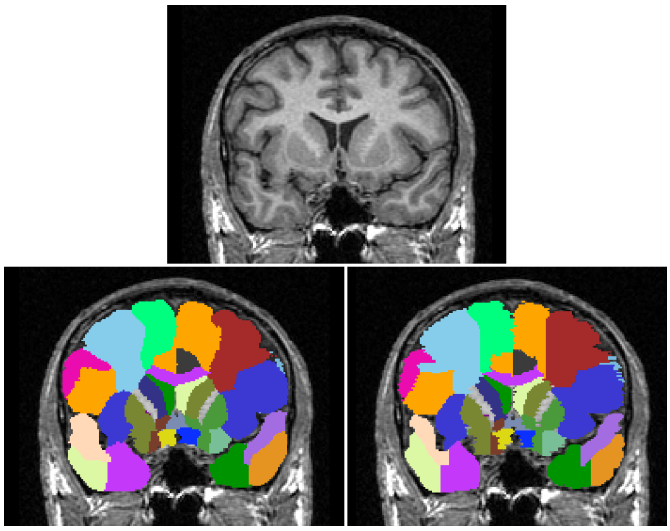


Figure 4.7: An example showing the template (top), the automated Multi-STEPS segmentation (bottom-left) and the manual segmentation (bottom-right). Note the smoothness of the boundaries for the automated segmentation method.

were kindly provided to us by the author. Results are shown in Table 4.5 and Figure 4.7.

Results show that STEPS with the MRF outperforms both STEPS without the MRF and MAPER in key internal structures, with 12 out of 83 structures showing a significant increase ($p < 10^{-4}$) in segmentation accuracy for the proposed method. The putamen was the only structure where MAPER outperformed STEPS, but that difference was not statistically significant.

The MRF introduced in this model not only results in a segmentation accuracy improvement but also improves the smoothness of the boundary between the labels. Anatomically, each one of the parcelated areas should have one single connected component. In order to test the advantages of introducing the MRF into the algorithm with regards to discontinuity, the average number of connected components per parcelated area and per subject was calculated for the proposed method with and without the MRF and for the manual . The average (SD) number of connected components per parcelated area was $8.9(1.3)$, $13.7(1.8)$, and $4.08(0.4)$ for STEPS with and without the MRF and for the manual respectively. A Welch's two-tailed paired t-test was performed in order to test significance. These results show a statistically significant ($p < 10^{-4}$) reduction in the mean number of connected components per parcelated

Table 4.5: Mean Dice coefficient for each structure, comparing the proposed method with and without MRF and MAPER (Heckemann et al., 2010). For bilateral structures, the left and right Dice score is averaged. Significantly higher ($p < 10^{-4}$) Dice scores are shown in bold font. Results for a set of key internal grey matter structures are shown.

	Fusion Method		
Structure Name	STEPS	STEPS no MRF	MAPER
Hippocampus	**0.842**	0.840	0.828
Amygdala	**0.805**	0.803	0.789
Caudate Nucleus	0.892	0.890	0.891
Nuc. Accumbens	**0.695**	0.687	0.682
Putamen	0.891	0.888	0.894
Thalamus	**0.894**	0.892	0.887
Globus pallidus	**0.798**	0.793	0.771

area when comparing STEPS with the MRF to STEPS without the MRF. STEPS with the MRF still performs significantly worse than the manual segmentation with regards to discontinuity and fragmentation of parcelated areas.

4.1.4 Discussion

In this section , I've developed an extension of the popular STAPLE algorithm that uses local intensity features to select the best labels to fuse, a novel iterative MRF to ensure spatial consistency and an uncertainty ROI optimisation to un-bias the algorithm towards larger structures. Both the robustness and accuracy of the segmentation were evaluated on the training set and in an independent database of cross-sectional and longitudinal brain MRI scans and tested the ability to directly use the segmentation for volumetric and atrophy rate measurements.

The algorithm was first tested on a simulated phantom with known ground truth segmentation, as a proof of concept. The STEPS method performed better than the STAPLE-GNCC method, presumably due to uncertainty caused by the lack of images in the template database with the same overall morphology as the image being segmented. Conversely, STEPS achieves a good overall segmentation due to the finer anatomical scale of the metric, suggesting that STEPS may enable the use of smaller template databases to describe the full population variability, leading to an improvement in both accuracy and computation time.

The proposed method was then applied to clinical data for the purpose of segmenting hippocampi. In order to find the parameters that produce the most accurate segmentations, a leave-one-out cross validation strategy was used to sample the overall accuracy from the parameter space. The proposed method was then compared to both STAPLE and majority voting under either a local or global label ranking strategy and also to the proposed method without the uncertainty ROI optimisation. Visual (see Fig. 4.2) and quantitative assessment demonstrates good segmentation accuracy and robustness, with the worst segmented image having a Dice score of 0.888. The proposed method performs significantly better ($p < 10^{-4}$) than all the other methods, demonstrating the advantage of combining the local ranking, the ROI optimisation strategies and the rater performance model. Furthermore, the proposed method obtains a (mean $\pm$ SD) Dice score (0.925 ± 0.021) close to the inter-rater variability of the manual segmentors (0.93 ± 0.03), assessed on a different database (Leung et al., 2010).

Another advantage of local ranking strategies is that they implicitly encode local morphological variability rather than global morphological variability. Fewer anatomical templates are needed to deal with the population's overall morphological variability. In order to test this idea, a second experiment was performed in order to show that local ranking can still obtain the higher segmentation accuracy as global ranking when using fewer anatomical templates. This is advantageous because if one can represent complex shapes with fewer samples, the need for a large and accurate template database is greatly reduced. The results of the jackknifing shown in section 4.1.3 demonstrate that STEPS can obtain significantly better segmentation accuracy, when measured using the Dice score, than the STAPLE-ROINCC label fusion algorithm, even when using three times fewer templates. As expected, there is a small shift of the Dice score peak for the optimal value of X between different database sizes with

the optimal value of X shifting to higher values with an increase in database size. Another interesting fact, not present in Fig. 4.5 is the consistent and significant reduction in the standard deviation of the Dice score per data set after the 10 simulations. This means that the proposed STEPS method not only produces better results but is also less dependent on the choice of data sets the template database is composed. This is important in situations where no knowledge is available about the morphology of a population or when the database size in inherently small.

This extra robustness with regards to database size can be exploited to improve computational efficiency. One can enforce morphological sparseness of the template database by learning the manifold structure of the data from a set of deformation fields to a group-wise space. This sparse representation of the morphological characteristics of the population would greatly reduce the computational complexity without degrading the segmentation accuracy. One should note that this effect was validated only on AD, MCI and controls using a template library based of AD and controls. It remains to be seen if results hold for hippocampi with different atrophy patterns and different intensity profiles such as in hippocampal sclerosis and certain atrophy syndromes like frontotemporal lobar degeneration and semantic dementia.

All experiments summarised above were performed on the training set using either leave-one-out cross validation or jackknifing. To test the performance of the fusing strategy on data from a different database acquired with a different MRI imaging systems and protocols, the same label fusion technique was also used to segment a subset of data from the ADNI database with manual segmentations. Using the parameters optimised in section 4.1.3, STEPS achieved a Dice accuracy above 0.9, significantly higher ($p < 10^{-4}$) than both ROINCC based fusion methods and also higher ($p = 0.006$) than the method proposed by Yushkevich et al. (2010). The performance improvement in STEPS when compared to Yushkevich et al. (2010) did not reach significance at the $p < 10^{-4}$ level due to the limited sample size, but an improvement trend is clearly visible. There was also an improvement in segmentation performance between STAPLE-LNCC and STEPS due to the addition of the uncertainty area ROI (as described in Section 4.1.2), showing the advantage of removing consensus areas from the parameter estimation. Furthermore, the flatter and larger plateau with high Dice scores when using STEPS makes the segmentation less sensitive to changes in the optimal parameters.

In a single label scenario, the STEPS algorithm was finally used to segment the hippocampi of all 682 1.5T ADNI data sets at baseline with 12-month repeat. Using the baseline data for a cross-sectional study, the volumetric results described in section 4.1.3 show the expected significant separability in terms of volume, between AD, MCI and controls. Using both the baseline and 12-month repeat in a longitudinal study, the results show again significant group discrimination between AD, MCI and controls. The atrophy rates are in line with those previously reported, with a mean hippocampal atrophy rate (%/year) of 4.04, 2.74 and 1.09 for the AD, MCI and control subjects respectively. These results were achieved using volumetric data from the binary hippocampal segmentations combined with tissue segmentation. I hypothesise that should baseline and followup scans be treated non-independently with regards to the template propagation or if the measurement of atrophy was changed to the boundary shift integral (Leung et al., 2010), our longitudinal measures would reduce in terms of noise or variability

with possibly improved disease group separation.

Lastly, in a multi-label propagation scenario, the algorithm was tested against MAPER (Heckemann et al., 2010). Results showed significant increases in segmentation performance, mainly in key internal grey matter structures like the hippocampus, amygdala, thalamus, globus pallidus and nucleus accumbens, known to be associated with several diseases. Furthermore, the statistically significant reduction in the number of connected components per structure shows the advantage of using STEPS with the MRF spatial smoothness term when compared to STEPS without MRF. Due to the locality of the similarity metric, I also speculate that the proposed methodology should provide improvements in the of pathological subjects and patients with different brain morphologies. However, further validation of multi-atlas based brain is necessary as the current findings are anecdotal for untested morphologies and pathological cases. This is specifically important in pathological situations that lead to large anatomical deformations (*e.g.* ventriculomegaly, highly atrophied brains), as some of these morphological changes might no be correctly captured by the non-rigid image registration step. In order to reduce the complexity and consequently the errors of the mapping between morphologically dissimilar images, the segmentations can be propagated via morphologically similar intermediate datasets using an approach similar to the one proposed by Wolz et al. (2010). The advantage of this kind of approach will be explored in section 4.2.

The current limitations of the proposed work are mostly related with the similarity metric. As previously described, even though the LNCC metric has many advantages when compared to a global metric, the local support of the metric can be problematic in low contrast areas. For example, if the non-rigid mapping between a normal subject and an AD patient with enlarged ventricles does not perform well enough, an area in the patient's ventricular cerebrospinal fluid can can be mapped and will correlate very well with the white matter area in the normal subject. This problem is caused by the local normalisation of the mean intensity between the two regions and can be ameliorated by a multi-level version of the same metric or by combining both local and global similarity metrics. Furthermore, as suggested by Souvenir and Pless (2007), the local intensity similarity metric can also be augmented by a morphological similarity metric based on the local displacement between mapped regions, thus introducing knowledge about anatomical shape changes. Nonetheless, the proposed framework is general enough and allows the replacement of the LNCC metric by any other similarity metric.

In this section, the focus has been on improving both the accuracy and robustness of segmentation propagation techniques by improving the label fusion component. Nonetheless, the algorithm's accuracy is still dependent, though to a smaller degree, on the quality of the manual segmentations and the type of pathologies and atrophy patterns represented in the template database. Further validation is still necessary in order to enable the unsupervised use of this algorithm in a clinical setting and for different disease groups. Additionally, the manual segmentation protocols can also be improved in order to avoid arbitrary cutoffs of structures, like the tail of the hippocampus, which may negatively affect the algorithm accuracy.

4.1.5 Conclusion

This section presents a new algorithm, called STEPS, that incorporates a fast locally normalised cross correlation (LNCC) based ranking combined with a consensus based ROI selection and a new iterative MRF into the STAPLE formulation. The algorithm was first tested on a database of manually segmented hippocampi using a leave-one-out cross validation. Results show a significant improvement in terms of Dice overlap when compared to state of the art GNCC and ROINCC based STAPLE and Majority Voting segmentations, achieving a mean Dice score of 0.925. The STEPS label fusion technique also achieved better accuracy than globally ranked techniques even when using only a third of the templates, diminishing the necessity of large template databases. When tested on an independent database with data sets from different MRI imaging systems and protocols, STEPS still achieved an average Dice score above 0.9, again significantly higher than globally ranked techniques. Furthermore, cross-sectional and longitudinal hippocampal volumetric studies showed expected significant differences in volume and atrophy rates between AD, MCI and controls. Finally, when applied to multi-atlas segmentation propagation, STEPS showed a statistically significant increase in segmentation accuracy in 12 (out of 83) brain structures when compared to MAPER.

4.2 GIF: Geodesic Information Flows

Homogenising the availability of manually generated information in large databases has been a key challenge of medical imaging for many years. Due to the time consuming nature of manually segmenting, parcellating and localising landmarks in medical images, these sources of information tend to be scarce and limited to small, and sometimes morphologically similar, subsets of data. In this section, I'll explore a new framework where these sources of information can be propagated to morphologically dissimilar images by diffusing and mapping the information through intermediate steps. The spatially variant data embedding uses the local morphology and intensity similarity between images to diffuse the information only between locally similar images. This framework can thus be used to propagate any information from any group of subject to every other subject in a database with great accuracy. Comparison to state-of-the-art propagation methods showed highly statistically significant ($p < 10^{-4}$) improvements in accuracy when propagating both structural parcellations and brain segmentations geodesically.

Derived publications and methodological sumary

- **Cardoso, M. J.**, Wolz, R., Modat, M., Rueckert, D., Ourselin, S.: Geodesic Information Flow. (2012) MICCAI

4.2.1 Methods

This section will first introduce the mathematical framework and the undirected graph for geodesic information flow, followed by the introduction of a distance metric that describes both image intensity similarity and morphological similarity between images. Finally, after building the graph embedding, the geodesic information propagation step is presented for two types of information: numerical and label attributes.

The implicit local data embeding

Let a set Y with R images be the full set of observed *grey-valued* anatomical data with the a-th image of this set denoted by $\mathbf{y}_a$. Each image $\mathbf{y}_a$ is going to be a vector of size N_a, with its j-th voxel denoted by y_{aj}.

In order to embed the observed data within a manifold, one normally starts by finding a distance between each pair of images. This distance provides insights about the global similarity between the images, and subsequently, about the manifold structure of the data. Theoretically, this global embedding assumes that one can represent the space spanned by the full data in a low dimensional space. However, due to the complexity of the data, the dimensions of the manifold can lack interpretability and usefulness. For example, Gerber et al. (2009) explored the manifold structure of the space of brain images and concluded that the first dimension of the manifold represents global ventricular expansion due to disease and ageing, while the second dimension meaning is described as "less obvious". Ideally, one would like to be able to capture the local manifold structure of the brain on a spatially constrained neighbourhood and not the global brain morphology. With one manifold representation per voxel, one would be able to describe the local brain morphology and similarity as a measure of distance between two mapped locations, and respective neighbourhoods, in two different images. However, due to computational and

memory requirements, this problem is untractable. As an example, only to store a pairwise distance matrix for one single image at every voxel and assuming a set of 120 neighbouring images with average size 200^3, one would need approximately 400GB of memory. Furthermore, the memory requirements will grow proportionally to R^2, where R is the number of datasets. Thus, one cannot have an explicit representation of the manifold at the voxel level. In this work, instead of constructing an explicit representation of the manifold, I implicitly represent the manifold through the local neighbourhood graph of each data point. This greatly reduces both computational complexity and memory requirements, making the problem tractable and linearly scalable with the number of datasets R in the database.

Let $\mathbf{D}$ be a set of distance matrices, with the a-th matrix of this set denoted by $\mathbf{D}_a$. Here, $\mathbf{D}_a$ will be an $N_a \times (R-1)$ matrix describing the distance between the image a and each one of the remaining $(R-1)$ images at every sample position j. More specifically, $d_{a\to b}(j)$ will contain the distance between the j-th sample of image $\mathbf{y}_a$ and its corresponding sample in image $\mathbf{y}_b$. I now introduce a threshold d_t over these distances. By doing so, one can now build an undirected graph where the neighbourhood of each data-point is restricted only to the data-points with $d_{a\to b}(j) < d_t$. Note that the graph is undirected only if the distances are a semi-metric (subadditivity is not required).

In this framework, one does not need to explicitly represent the full graph. In order to solve the information diffusion problem at a given location, one only needs to keep track of the graph neighbourhood at that specific location, visually shown in Fig. 4.8 - right. Here, the realm of observations (the blue connections) from the data point in bold is limited by its direct neighbouring nodes with distances below d_t. Under this undirected graph assumption, a heat kernel decay function $w_{a\to b}(j)$ is then used to diffuse the information (Kondor and Lafferty, 2002). This kernel is defined as

$$w_{a\to b}(j) = \begin{cases} e^{-\frac{d_{a\to b}(j)}{t}} & d_{a\to b}(j) < d_t \\ 0 & else \end{cases} \tag{4.14}$$

with t being a heat kernel temperature that will determine the speed and the distance the information can diffuse. In this work, $d_t = t$, meaning that the choice of t will determine both the maximum distance traveled by the information and the amount of information diffusion that occurs at each iteration.

The distance metric

The heat kernel decay function is based on the assumption that one can calculate a distance between two nodes in the graph. This distance should be at least a semi-metric, respecting both the coincidence and separation axioms and symmetry. In a medical imaging framework, and more specifically in neuroimaging, the local distance between images should take into account both local morphology and local image similarity. To achieve this goal, Gerber et al. (2009) propose the use of coordinate transformations as a distance metric that informs about the object's morphology. These coordinate transformations map an image $\mathbf{y}_a$ to an image $\mathbf{y}_b$ by finding the optimal transformation $\mathbf{t}_{a\to b}$ that minimises some cost function. In order to be a semi-metric, this coordinate transformation has to be symmetric, inverse consistent and diffeomorphic. Here, a symmetric variant of a non-rigid free-form registration algorithm is used (Modat et al., 2010). Under the symmetry and diffeomorphism constraints, the transformation $\mathbf{t}_{a\to b} = \mathbf{t}_{b\to a}^{-1}$ and

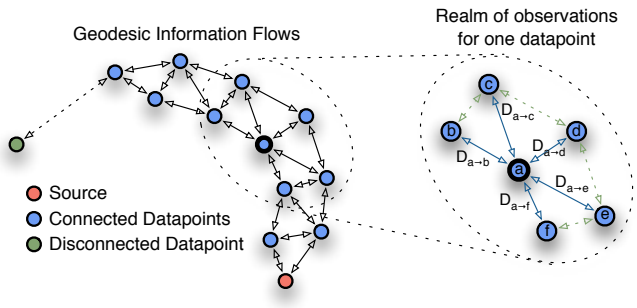


Figure 4.8: Left) Implicit manifold with the neighbourhood defined as all the data points within a certain distance. Note that if the manifold is sparse, some data points might be disconnected. Right) Diagram representing the observed and unobserved connections (in blue and green respectively) and distances from the standpoint of the data point a.

$\mathbf{t}_{a\to b} \circ \mathbf{t}_{b\to a} = \text{Id}$, with $\mathbf{t}^{-1}$ being the inverse of the transformation, $\circ$ being the composition operator and Id the identity transformation. In order to remove the smoothly varying local affine component of the transformation that characterises the global anatomical shape differences, the low frequency component of the transformation is removed. From the remaining high-frequency version of the transformation, one can then find the displacement field $\mathbf{f}_{a\to b}$ that describes how much a point j in $\mathbf{y}_a$ had to move in order to match the corresponding point j in $\mathbf{y}_b$.

Even though this displacement field will describe the morphological differences between the brains, I also combine it with an intensity similarity metric in order to assess the local similarity between the images after transformation (Souvenir and Pless, 2007). This similarity term is necessary to characterise both the local differences in tissue appearance due to pathology (*e.g.* damaged WM in dementia) and also some possible local registration errors. The local similarity between an image $\mathbf{y}_a$ and an image $\mathbf{y}_b$ transformed by $\mathbf{t}_{b\to a}$, denoted by $\mathbf{l}_{a\to b}$, can be calculated as the local sum of squared differences (LSSD) between the intensity in these images, using a cubic B-spline kernel as a local smoothing function. The two metrics are combined together by setting $d_{a\to b}(j) = \alpha l_{a\to b}(j) + (1-\alpha) f_{a\to b}(j)$, with α being a relative weight, meaning that both a low displacement and a low LSSD are necessary to obtain a low distance $d_{a\to b}(j)$ between images. In this work $\alpha = 0.5$. Optimisation of α will be part of future work.

Geodesic Information Flows

The two previous sub-sections have defined the neighbourhood graph and the distance metric. This sub-section will make use of the graph structure to introduce the concept of propagating information between neighbouring nodes of the graph.

Let I be a set of vectors, with the a-th vector of this set denoted by $\mathbf{i}_a$. Each vector $\mathbf{i}_a$ has its j-th element denoted by i_{aj}, representing the associated information at location j for image $\mathbf{y}_a$. Assuming that the information to be propagated is only available in a subset of images within the full database, only some of the vectors $\mathbf{i}_a$ will be defined. Now, let an indicator vector $\mathbf{k}_a$, indexed by k_{aj}, characterise the status of the information at location i_{aj}. Here, k_{aj} is set to 2 if i_{aj} is a source of information, set to 1 if the information has diffused and reached the location i_{aj} or 0 if the information does not exist at

location i_{aj}. As the realm of observations at each spatial location is limited by its closest neighbours, the best approximation for data flow is given by a normalised weighted sum of the information available within the neighbourhood. Thus,

$$i_{aj} = \frac{\sum_{\forall b \in k_{bj}>0} w_{a\to b}(j) i_{bj}}{\sum_{\forall b \in k_{bj}>0} w_{a\to b}(j)} \tag{4.15}$$

with j being the spatially transformed coordinate j into the space of image $\mathbf{i}_b$, mapped using the previously described transformation. The information flow is thus governed by the heat kernel-derived weights $w_{a\to b}(j)$. If the set $b \in (k_{bj} > 0)$ is not empty, the information in i_{bj} can then be propagated to i_{aj}. In this case, k_{aj} is set to 1, else k_{aj} is set to 0. Note that Eq. 4.15 is only valid for floating point data propagation like a probabilistic atlas. The same equation can be re-formulated in a weighted label fusion scheme, by making i_{aj} equal to $p(i_{aj}, l)$, representing the probability that location j in image a has label l, and by making i_{bj} equal to $p(i_{bj}, l)$, representing the probability that location j has label l in image b. Eq 4.15 is solved iteratively for all samples where $k_{aj} < 2$, until all are marked as solved, *i.e.* $k_{aj} = 1$. The number of times i_{aj} could not be solved because $k_{bj} > 0 \; \forall \; b$ is an empty set, represents the number of steps through the graph's geodesic path necessary to transport the information from its source, which is related to the amount of extrapolation of information and propagation accuracy.

Finally, because the temperature t will determine the neighbourhood size and consequently the existence of a connection from every information source to all the targets, the geodesic information flow is solved multiple times for several values of t. This temperature t is varied between 5 and 1 with decrements of 0.5. In a similar fashion to an annealing process, the information at the lowest possible temperature is kept as the answer, as the distance traveled by each extrapolation step is the lowest.

4.2.2 Validation

The data used in this validation section, depicted in Fig. 4.9, is compromised of two sets: 30 young controls with associated structural parcellation of 83 key structures Hammers et al. (2007) (`http://www.brain-development.org`); 90 subjects from the ADNI database (`http://adni.loni.ucla.edu`), subdivided into 30 controls, 30 Mild Cognitive Impairment (MCI) and 30 Alzheimer's diseased (AD) patients with associated manual segmentations of the brain. The current work aims at homogenising databases under the assumption that extra information is only available on a subset of the data. From these sources of information, measuring the information extrapolation accuracy will always be limited by the anatomical and pathological variability within the full dataset and by the range of available segmentations. Furthermore, the most complex sources of information, like the 30 young controls with full brain parcellations, are simply not available in pathological subjects. This makes the validation anecdotal for untested morphologies. The proposed validation will thus have two components. First, the overlap accuracy of multi-label information propagation will be estimated and compared to MAPER (Heckemann et al., 2010) using a leave one out approach on the 30 young controls. Then, the accuracy of information extrapolation will be characterised by propagating the brain segmentations from the elderly control group to the MCI and AD patients.

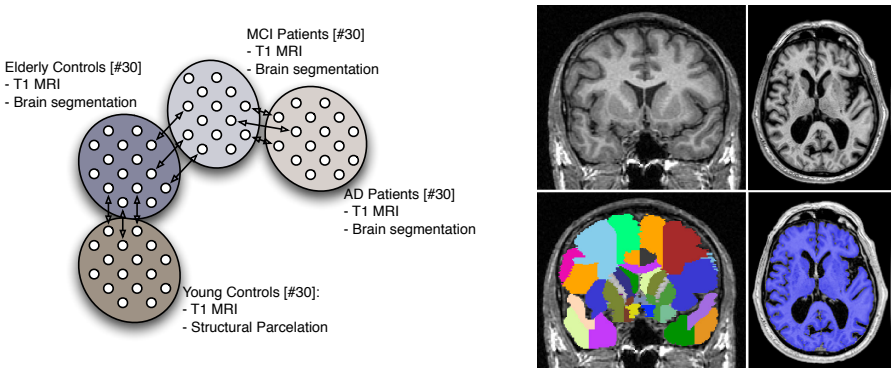


Figure 4.9: (Left) All the sets of data used in this work and their associated information. (Right) One dataset from the Hammers atlas with its associated structural parcellation below; an ADNI AD subject with the associated brain segmentation

Table 4.6: Mean Dice coefficient for a set of key structures, comparing the proposed method (GIF) with MAPER Heckemann et al. (2010). Statisticaly higher mean Dice is shown in bold font. Only a limited number of structures are shown due to restrictive space availability.

Structure	Unilateral Structures		
	GIF	MAPER	p-value
All Structures	0.8179	0.8089	$< 10^{-4}$
Corp. callos.	0.8802	0.8674	$< 10^{-4}$
Brainstem	0.9534	0.9377	$< 10^{-4}$

Structure	Left Side			Right Side		
	GIF	MAPER	p-value	GIF	MAPER	p-value
Hippocampus	0.8439	0.8335	0.0048	0.8251	0.8211	0.2036
Amygdala	0.8263	0.7922	$< 10^{-4}$	0.8241	0.7830	$< 10^{-4}$
Caudate nucl.	0.8983	0.8923	0.0380	0.9004	0.8955	0.0478
Nucleus acc.	0.7581	0.6834	$< 10^{-4}$	0.7317	0.6707	$< 10^{-4}$
Putamen	0.9069	0.8916	$< 10^{-4}$	0.9107	0.8959	$< 10^{-4}$
Thalamus	0.9210	0.8879	$< 10^{-4}$	0.9205	0.8852	$< 10^{-4}$
Pallidum	0.8562	0.7661	$< 10^{-4}$	0.8554	0.7672	$< 10^{-4}$

Multi-label propagation accuracy

The accuracy of propagating information through a geodesic path was compared to MAPER (Heckemann et al., 2010), a direct information fusion method based on majority voting. As the amount of parcellations available for validation is limited, a leave-one-out cross validation was performed only on the 30 young controls that have manual brain parcellations. One should note that the limited availability of segmentations restricts the range of morphological variability in the propagation, thus not representing the real performance when segmenting morphologically dissimilar subjects.

The Dice score was used as a measure of accuracy. The mean Dice scores per structure for the the leave-one-out cross validation are shown in Table 4.6. Out of 83 structures, 15 structures had a significantly higher Dice score using the Geodesic information Flow when compared to MAPER, while only two structures (lingual gyrus and superior parietal gyrus) where better segmented in MAPER. The mean Dice score over all structures and all patients for the proposed method (0.8197) was significantly higher ($p < 10^{-4}$) than in MAPER (0.8089). An example of the propagation to a highly atrophied subject from the ADNI database is shown in Fig. 4.2 (right).

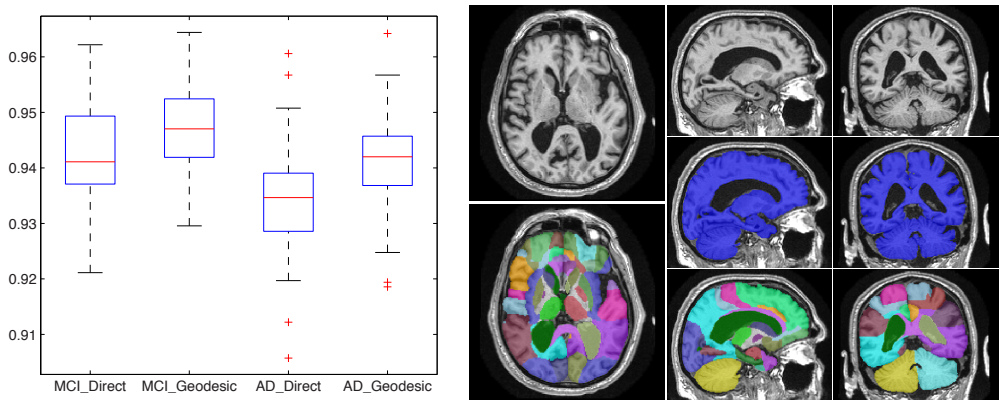


Figure 4.10: (Left) Dice scores for direct and geodesic propagation of brain mask. (Right) An example of the propagation of both the structural parcellation and brain segmentation to an highly atrophied AD subjects (ID:1281) from the ADNI database. Note the correct ventricle segmentation and the smooth deep grey matter parcellation.

Information Extrapolation Accuracy

In the previous sub-section, the accuracy of propagating information through a geodesic path was limited to a morphologically similar set of subjects. Thus, the previous validation will not capture the ability to extrapolate information to anatomically disparate subjects. The information extrapolation accuracy is thus assessed by using only a subset (the elderly control group) of all the manual brain segmentations. This morphologically clustered set of data is then used to segment both the MCI and AD groups. The proposed geodesic propagation algorithm is compared to a direct propagation algorithm based on the locally weighted majority voting algorithm with a inverse exponential weight proposed by Yushkevich et al. (2010). The results are presented in Fig. 4.10 (left), with segmentation accuracy measured using Dice similarity. The mean (std) Dice score for the proposed geodesic method was 0.940(0.009) and 0.947(0.008) for the AD an MCI groups respectively while for the direct method, the mean (std) Dice score was 0.934(0.009) and 0.942(0.008) for the AD an MCI groups respectively. This represents a statistically significant ($p < 10^{-4}$) increase in segmentation accuracy when using a two-tailed paired t-test for statistical comparison. Note that one should not compare these results with other brain segmentation methods due to the lack of post-processing and the limited size of the training set.

4.2.3 Conclusion

This section presents an algorithm where information is geodesically propagated through a local implicit neighbourhood graph. Application to structural parcellation and brain segmentation propagation has demonstrated the significant ($p < 10^{-4}$) advantages of the proposed framework when compared to state of the art methods. Overall, the proposed framework can be used to propagate any information from a group of subjects to every other subject in a dataset.

Chapter 5

Cortical Thickness Estimation

5.1 KaTE: Topologically Correct Thickness Measurements using Khalimsky's Cubic Complex

The extraction of thickness measurements from shapes with spherical topology has been an active area of research in medical imaging. Measuring the thickness of structures from automatic probabilistic segmentations is normally hindered by the presence of noise, PV effects and the limited resolution of medical images. Also, the complexity of certain shapes, like the highly convoluted and PV corrupted cerebral cortex, results in topologically inconsistent measurements. In this section I explore the use of Khalimsky's cubic complex for the extraction of topologically correct thickness measurements from probabilistic or fuzzy segmentations without explicit parametrisation of the edge. A sequence of element collapse operations is used to correct the topology of the segmentation. The Laplace equation is then solved between multiple equipotential lines and the thickness measured with an ordered upwind differencing method using an anisotropic grid with the probabilistic segmentation as a speed function. Experiments performed on digital phantoms show that the proposed method obtains topologically correct thickness measurements with an increase in accuracy when compared to two well established techniques. Furthermore, quantitative analysis on brain MRI data showed that the proposed algorithm is able to retrieve expected group differences between the cortical thickness of AD patients and controls with high statistical significance.

5.1.1 Derived publications and methodological sumary

- **Cardoso, M. J**., Clarkson, M. J., Modat, M., Ourselin, S., On the Extraction of Topologically Correct Thickness Measurements using Khalimsky's Cubic Complex, (2011), IPMI

- **Cardoso, M. J**., Clarkson, M. J. , Ridgway, G. R., Modat, M. , Talbot, H., Couprie, M., Ourselin, S., Topologically correct cortical segmentation using Khalimsky's cubic complex framework, (2010), SPIE

The aim of this work is to combine the features of all the above voxel-based methods in a unified, fully automated **Kha**limsky based **T**hickness **E**stimation algorithm, called **KaTE**, that is topologically correct and partial-volume aware. I use the properties of the Khalimsky grid and an iterative set of ele-

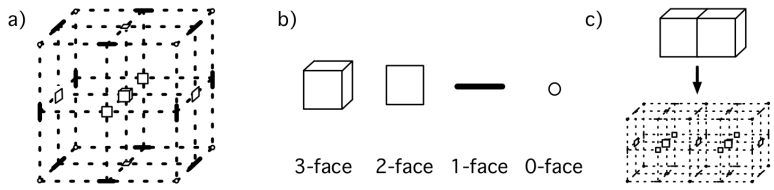


Figure 5.1: Example of: a) a Cubic Complex, b) its m-face and c) the object in the cubical space associated with neighbouring voxels

ment collapse operations to correct the topology of the segmentation. I then extend the single Laplacian field method to a multi-stage Laplacian in order to encompass the partial-volume containing areas and I modify the classic PDE proposed by (Yezzi Jr and Prince, 2003) to have a spatially varying speed function that is dependent on the probabilistic segmentation.

5.1.2 Methods

Topology preservation and the Khalimsky's Cubic Complex

Topology-preserving operations are used in many image analysis applications in order to transform an object while leaving its topological characteristics unchanged. A typical topology-preserving transformation based on simple point deletion can be described as follows. Let the object be a set X of points in the voxel grid, and a subset K of X be a constraint set. At each iteration, choose a simple point x_i in X_0 but that is not in K according to a priority function (e.g, Euclidean distance) and set $X_{i+1} = X_i \backslash x_i$. Iterate until no other point can be removed. This process results in a homotopic skeleton of X constrained by K. Notwithstanding the simplicity, thinning algorithms that work in the voxel space have some well-described problems regarding the minimality of the set X_n and the existence of lumps (Passat et al., 2008; Cointepas et al., 2001). Abstract complexes, like the Khalimsky space (Khalimsky et al., 1990) have been promoted, in particular by V. Kovalevsky (Kovalevsky, 1989), to provide a sound topological basis for image analysis. Intuitively, a cubic complex can be seen as a space where every voxel is represented by a structure composed of a set of elements having various dimensions (*e.g.* cubes, squares, edges, vertices) put together according to some rules. For illustrative purposes see Fig. 5.1. Let $\mathbb{Z}^3$ denote the original 3D cartesian voxel grid and $\mathbb{F}^3$ the set composed of all m-faces of $\mathbb{Z}^3$ for all $m = 0$ to 3. An m-face of $\mathbb{Z}^n$ is called a point if $m = 0$, a line $m = 1$, a square if $m = 2$ and a cube if $m = 3$. Please refer to (Passat et al., 2008; Cointepas et al., 2001) for a complete formal description of the cubic complex framework. For our application, and as a voxel is equivalent to an elementary cube, an easy correspondence can be made between the voxel classical view and the framework of cubical complexes. In the rest of the section, I use the term voxel to mean a 3-cell. This abstract space provides a sound basis for digital topology and topology correction but also, due to the dimensionality of its elements, an interesting framework for thickness measurement extraction. Within this framework, one is able to connect different areas with infinitesimally thin surfaces, correcting the topology of the segmentations with minimal influence on the thickness estimation step.

Collapse Operation and Topology Correction

Collapse operation: The collapse operation is a well-known operation that preserves topology. Let F be a complex in $\mathbb{F}^3$ and let $f \in F$. If g is a *proper face* of F and f is the only face of F which *strictly includes* g, then the pair (f, g) is considered a free pair of F. Here, the dimension of the face g is $dim(f) - 1$. As the pair (f, g) is a free pair, it can be removed without any change in the topology of F by an operation called *elementary collapse* defined as $F \backslash \{f, g\}$. This operation is the $\mathbb{F}^3$ equivalent to the removal of a simple point in $\mathbb{Z}^3$.

This operation can then be used to collapse a shape into another while maintaining the same topology. In a more general form, let the operation that collapses the set F onto the constraint set K, with a priority function P be denoted by $\text{Col}(F, K, P)$, where a lower value of P corresponds to a higher priority.

Topology correction of spherical objects: Starting from a probabilistic segmentation of a spherical object, let S_{In}, S_{Out} and S_{Obj} be probabilistic segmentations of the internal area, external area and the object of interest. Thus, S_{In_i}, S_{Out_i} and S_{Obj_i} represents the probability for voxel i in $\mathbb{Z}^3$ to belong to each corresponding class. These probabilistic segmentations can be converted from $\mathbb{Z}^3$ to $\mathbb{F}^3$ by setting the value at each m-face equal to the average of their $m + 1$ surrounding faces, for all $m \in 0, 1, 2$ and the value at the 3-face as the original value for the corresponding voxel in $\mathbb{Z}^3$. For example, the value at the 0-face, $S_{Obj_{im}}$ with $m = 0$ will be equal to the average of its surrounding 1-face. This is equivalent to an ordered linear interpolation.

The topology correction for spherical objects will be done in four steps: in step 1, starting from an object with an Euler characteristic of 1, in this case a box bounding the object in $\mathbb{F}^3$ defined as $F_{Box_{im}}$, set $F_{\text{In}_{\text{pv}}} = \text{Col}(F_{\text{Box}}, K_{\text{In}_{\text{pv}}}, S_{\text{In}})$ with the constraint set $K_{\text{In}_{\text{pv}}} = (S_{\text{In}} > t)^-$ as the topological closure (represented by the $^-$ operation) of all the elements of $S_{\text{In}_{im}}$ above a certain threshold t. This constraint set contains all the voxels that have more that t probability of belonging to the inner area. On an ideal segmentation, the value of threshold t would be 0, however, due to noise, t is set to 0.05 for probabilistic segmentation on standard clinical datasets. This operation is the Khalimsky equivalent of shrink wrapping the constraint set without a change in topology, resulting in a set that has all the voxels that have a probability above t of belonging to the inner area while maintaining its initial Euler characteristic of 1.

In step 2, set $F_{\text{In}_{\text{pure}}} = \text{Col}(F_{\text{In}_{\text{pv}}}, K_{\text{In}_{\text{pure}}}, S_{\text{In}})$ with $K_{\text{In}_{\text{pure}}} = (S_{\text{In}} > (1 - t))^-$ as the constraint set. This will continue the collapsing of the previous structure $F_{\text{In}_{\text{pv}}}$ ordered by the probability of belonging to the inner area S_{In} and will result in a set that contains all the voxels that have a probability above $(1 - t)$ of belonging to the inner area while maintaining its initial Euler characteristic of 1.

In order to get a topologically corrected $F_{\text{Out}_{\text{pv}}}$, the set $F_{\text{In}_{\text{pure}}}$ has to be inverted. However, in a cubic complex space, the inversion of a complex is not a complex. A topological closure could be used to transform the inverted set into a complex, but this operation would change the topological characteristics of the inverted set. Under the assumption of smoothness for the probabilities S and consequently smoothness of the edge of $F_{\text{In}_{\text{pure}}}$, one can remove all the *incomplete* m-face from the complex and still

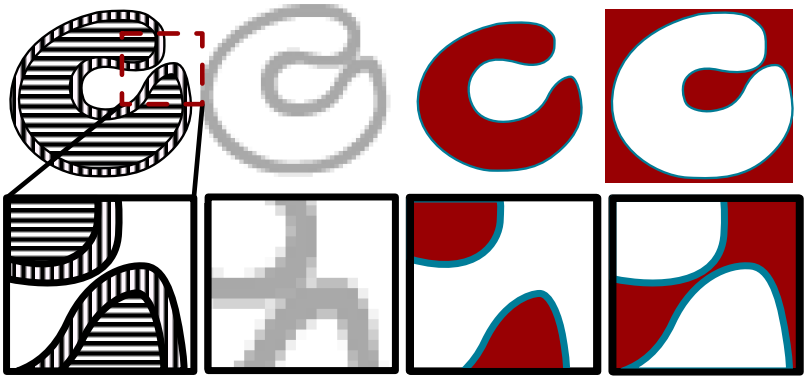


Figure 5.2: From left to right: An example object with the outside in white, the inside in horizontal lines and the object of interest with vertical lines; A rasterised version of the object with PV effect (notice the erroneously connected structure due to PV); the result of the bounding box collapse after stage 1 in blue and after stage 2 in red; the result of the collapse of the inverse of the topology corrected pure inside area after stage 3 in blue and after stage 4 in red;

leave the topological characteristics unchanged. This operation consists of removing all the external m-faces that are not limited by all their n-face elements, with $n < m$, in an ordered manner from $m = 3$ to 0. For example, a 2-face (square) will be removed if any of it's surrounding 1-faces (lines) and 0-faces (points) do not belong to the inverted set. This operation can be described as an opening on the incomplete m-faces of the set.

Let the inverted set of $F_{\text{In}_{\text{pure}}}$ according to the above described method be represented by $F^{-1}_{\text{In}_{\text{pure}}}$. Then set $F_{\text{Out}_{\text{pv}}} = \text{Col}(F^{-1}_{\text{In}_{\text{pure}}}, K_{\text{Out}_{\text{pv}}}, S_{\text{Out}})$ with the constraint set $K_{\text{Out}_{\text{pv}}} = (S_{\text{In}} > t)^{-}$. This will collapse the inverted set $F^{-1}_{\text{In}_{\text{pure}}}$ ordered by the probability of belonging to the outer area (S_{Out}) and will result in a set that contains all the voxels that have a probability above t of belonging to the outer area. Due to the inversion, and assuming that the foreground object is limited by the image boundaries, it's Euler characteristic will now be equal to 2 (spherical topology).

Finally, in step 4, set $F_{\text{Out}_{\text{pure}}} = \text{Col}(F_{\text{Out}_{\text{pv}}}, K_{\text{Out}_{\text{pure}}}, S_{\text{Out}})$ with the constraint set $K_{\text{Out}_{\text{pure}}} = (S_{\text{Out}} > (1-t))^{-}$. This will continue the collapsing of the previous structure $F_{\text{Out}_{\text{pv}}}$ ordered by the probability of belonging to the outer area (S_{Out}) and will result in a set that contains all the voxels that have a probability above $(1-t)$ of belonging to the outer area while maintaining its previous Euler characteristic of 2.

In short, this sequence of operations will produce a $F_{\text{In}_{\text{pure}}}$ and a $F_{\text{In}_{\text{pv}}}$ set with a Euler characteristic of 1 and a $F_{\text{Out}_{\text{pure}}}$ and a $F_{\text{Out}_{\text{pv}}}$ set with a Euler characteristic of 2. Please refer to Fig. 5.2 for a visual representation of the results of the pipeline.

Apart from the minimality of the skeletons and the existence of lumps, the cubic complex has several other benefits over voxel based topology correction as the same sequence of steps will result in incorrect topologies when the thickness of the structure of interest is close to the image resolution. The problem is represented visually in Fig. 5.3.

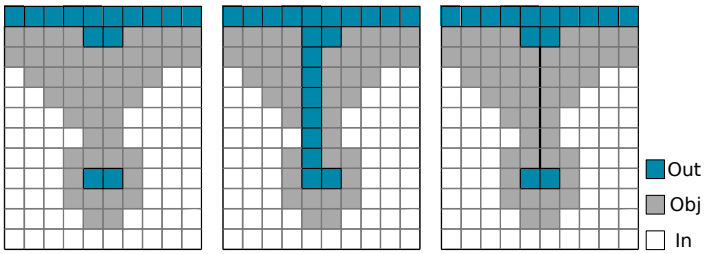


Figure 5.3: Left: Disconnected $K_{\text{Out}_{\text{pure}}}$, Centre: Voxel based skeletonisation of $S_{\text{Out}} \bigcup S_{\text{Obj}}$, constrained by K_{Out}. Right: The same colapse procedure in the Khalimsky space. Notice the change of topology of the S_{Obj} structure when using a normal skeleton.

Multi-stage Laplace equation and Thickness measurement

In order to calculate thickness, a unique association between two points is required. I use the Laplace equation, a second order partial differential equation (PDE), solved between two enclosed boundaries Ω and Ω'. The classic equation takes the form

$$\nabla^2 \phi = 0 \tag{5.1}$$

with the value at ϕ_Ω and $\phi_{\Omega'}$ set up as boundary conditions. For this work, instead of a single Laplacian field for all the pure voxels as in (Hutton et al., 2008; Jones et al., 2000; Yezzi Jr and Prince, 2003; Rocha et al., 2005; Acosta et al., 2009), a multiple Laplacian field is solved. The idea behind this approach came from the fact that methods that integrate through PV using ray casting (Acosta et al., 2009) have problems estimating surface normals correctly. The use of a topologically correct extended Laplacian field implicitly solves this problem as the direction of integration becomes known.

The first stage, represented in Fig. 5.4 as a light blue colour, is solved between the $F_{\text{In}_{\text{pure}}}$ and the $F_{\text{In}_{\text{pv}}}$ edges, where the outer $(0, 1, 2)$-face of the respective complexes are fixed and set to 0 and 100. The second stage, represented in Fig. 5.4 as a white colour, is solved between the $F_{\text{In}_{\text{pv}}}$ and the $F_{\text{Out}_{\text{pv}}}$ edges, where the outer $(0, 1, 2)$-face of the respective complexes are set to 100 and 200. Finally, in the third stage, represented in Fig. 5.4 as a light grey colour, is solved between the $F_{\text{Out}_{\text{pv}}}$ and the $F_{\text{Out}_{\text{pure}}}$ edges, where the outer $(0, 1, 2)$-face of the respective complexes are set to 200 and 300. By solving the Laplace

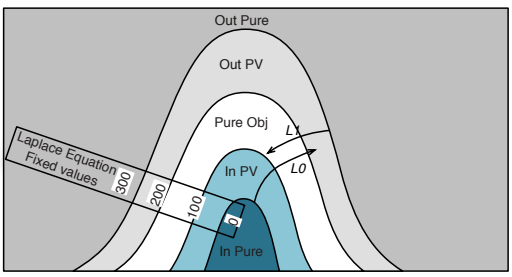


Figure 5.4: Multi-Stage Laplace equation: The lines between each class are set to a fixed potential and the Laplace equation is solved. The distance L_0 and L_1 is integrated from opposite sides of the object, following the perpendicular to Laplacian field isolines.

equation between these four equipotential lines, a smooth transition field arises. Note that the pattern of streamlines and isolines is independent of the choice of boundary condition voltages, as long as the four voltages, here set to 0, 100, 200 and 300, are different, equally spaced and increasing. Considering that the original image might have an anisotropic voxel grid, and assuming that the value of the Laplacian in each element is the value at the centre of mass of the element, the finite difference approximation can be seen as:

$$\phi_{i+1}(x,y,z) = \frac{s_y^2 s_z^2 D(x \pm s_x, y, z) + s_x^2 s_z^2 D(x, y \pm s_y, z) + s_x^2 s_y^2 D(x, y, z \pm s_z))}{2(s_y^2 s_z^2 + s_x^2 s_z^2 + s_x^2 s_y^2)} \tag{5.2}$$

with $D(x \pm s_x, y, z) = \phi_i(x + s_x, y, z) + \phi_i(x - s_x, y, z)$ and similarly for the y and z directions. Here s_x, s_y and s_z equal to the distance between the centre of mass of neighbouring m-faces in the x, y and z respectively. In this case, s_x, s_y and s_z will be equal to half the original voxel size in the x, y and z directions on the $\mathbb{Z}^3$ image. From the resultant Laplacian field, the normals to the direction of the Laplacian isolines, denoted by T, are calculated using finite differences.

In order to measure thickness, the length of the streamlines between the inner and outer surface has to be measured by integrating the vector field T. This can be achieved by the use of two first order linear PDEs. Instead of the basic form proposed by Yezzi (Yezzi Jr and Prince, 2003), where the speed of the advancing front is assumed to be 1, I use a more generalised form of the PDE. Here,

$$\nabla L \cdot T = f \tag{5.3}$$

for an unknown function L and assuming that T and f are known. In our case, and differently from (Jones et al., 2000; Yezzi Jr and Prince, 2003; Rocha et al., 2005; Acosta et al., 2009), the value of f will be spatially varying and equal to the probability of belonging to the object, $f = S_{\text{Obj}}$. This value will act as *time cost* and will make the value of L equivalent to the time of arrival in a level-set framework. Even though collisions of the advancing front might exist, they are not a problem due to the upwind nature of the integration and the existence of the vector field T. Let $L_{0_{(x,y,z)}}$ be a function that measures the time of arrival (arc length of the streamline according to the time cost f) from the boundary of set $F_{\text{In}_{\text{pure}}}$ to a point in the object and $L_{1_{(x,y,z)}}$ be the time of arrival from the boundary of set $F_{\text{Out}_{\text{pure}}}$ to the point in the object. Thus, similarly to (Yezzi Jr and Prince, 2003),

$$\begin{aligned} L_{0_{(x,y,z)}} = \frac{1}{(a_y a_z |T_x| + a_x a_z |T_y| + a_x a_y |T_z|)} \big[& f_{(x,y,z)} a_x a_y a_z + \\ & a_y a_z |T_x| * L_{0_{(x \mp a_x, y, z)}} + a_x a_z |T_y| * L_{0_{(x, y \mp a_y, z)}} + \\ & a_x a_y |T_z| * L_{0_{(x,y,z \mp a_z)}} \big] \end{aligned} \tag{5.4}$$

and equivalently for L_1 by replacing the $\mp$ with $\pm$. Here, a_x, a_y and a_z are the size of the m-face in the x,y,z dimension respectively, and $f_{(x,y,z)} = S_{\text{Obj}}(x,y,z)$. The notation $x \pm a_x$ and $x \mp a_x$ and similarly for y,z,a_y and a_z are defined as

$$x \pm a_x = \begin{cases} x + a_x & |T_x| > 0 \\ x - a_x & |T_x| < 0 \end{cases} \qquad x \mp a_x = \begin{cases} x - a_x & |T_x| > 0 \\ x + a_x & |T_x| < 0 \end{cases} \tag{5.5}$$

As some elements are infinitesimally thin and this would cause numerical instability, instead of the real structure size, a will be defined as the size of an influence zone. The influence area size will be equal to αs, with $\alpha \in]0, 1]$, in the infinitesimally thin direction of an m-face and equal to $(2-\alpha)s$ in the other directions. For example, a 2-face in the xy plane will have $a_x = (2-\alpha)s_x$, $a_y = (2-\alpha)s_y$ and $a_z = \alpha s_z$. If $\alpha = 1$, the above equation will reduce to the classic anisotropic form and the behaviour will be equivalent to considering each element as a normal $\mathbb{Z}^3$ voxel, while α close to 0 will be the equivalent to having an infinitesimally thin element for every m-face with $m < 3$. Please refer to Fig. 5.5 for a visual explanation of the influence areas in a 2D complex.

The final value of thickness is then defined as $Thick = L_0 + L_1$. In order to un-bias any further statistical analysis, the value of thickness should only be calculated either at the mid-harmonic point within the object, where $L_0 = L_1$, or at one of the mid-isopotential lines. The ribbon containing the thickness measurements will thus have spherical topology (Euler characteristic of 2).

5.1.3 Experiments and Results

This experimental section is divided into two sections. First, the performance of the proposed algorithm is assessed against a digital phantom with known ground truth thickness and the results compared to two state-of-the-art methods. The proposed method is then applied to brain MRI data in order to assess group separation in terms of cortical thickness between Alzheimer's disease diagnosed patients and controls.

Phantom validation

In order to evaluate if the proposed method can accurately retrieve the underlying thickness of an object, 6 folded 3D digital phantoms with spherical topology and known ground truth thickness were created (Fig. 5.6), resulting in six $300 \times 250 \times 250$ isotropic images with 3 structures - In, Out and Object. The thickness of the object is changing with time from 4.8 to $3.6mm$ in order to simulate a thickness lost in an object over several time points. These high resolution phantoms were then down-sampled by 5 in order to simulate PV effect and the thickness of the down-sampled structures was then measured. I compare the proposed method with the classic Jones et al. (2000) method that works on a binary version of the segmentation. Here, two threshold values (0.5 and 0.9) were used to binarise the data. The method proposed by Acosta et al. (2009) was also used for this comparison as it also integrates through PV using

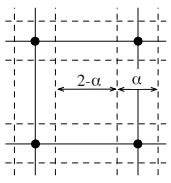


Figure 5.5: A 2D complex represented by the full lines and the points. The influence area represented in dashed lines and determined by the value of α.

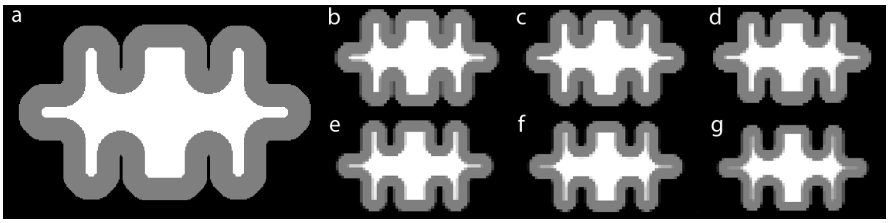


Figure 5.6: a) A slice of the 3D simulated high resolution phantom with $4.8mm$ thickness. b) The same slice in the down-sampled version of the phantom with thickness 4.8, c) 4.6, d) 4.4, e) 4.2, f) 4.0 and g) $3.6mm$, equivalent to a thinning procedure of $0.4mm$ per year, sampled at year 0, 0.5, 1, 1.5, 2 and 3.

a ray casting technique.

Results show that the method developed by Jones et al. (2000) is very sensitive to threshold selection, mainly because there is no topological information when extracting the boundaries of the object of interest. In real world scenarios, the selection of a threshold becomes more critical as the segmentation is less well behaved. Thus, methods that require threshold selection should not be used without optimisation of the threshold parameter. Furthermore, this optimisation requires a ground truth, which is not available on clinical data, and has to be done per dataset, hampering the utilisation of this technique. The method proposed by Acosta et al. is not dependant on a threshold value as it integrates through PV using a ray casting technique. However, the choice of stopping criteria for the ray casting is problematic in convoluted areas, greatly increasing the standard deviation of the estimated thickness. When applied to real data, this increase in standard deviation will reduce the statistical power for group analysis. Using a one-sample t-test, none of the methods estimate a thickness that statistically differs from the ground truth. However, the proposed method achieves better accuracy and precision than the other two methods. Furthermore, the proposed method is the only one that guarantees spherical topology of the estimated thickness. The pattern of difference between ground truth and estimated thickness is represented in Fig. 5.7(c,d). This difference is, in overall, very close to zero for KaTE but very variable for Acosta *et al.*(Acosta et al., 2009).

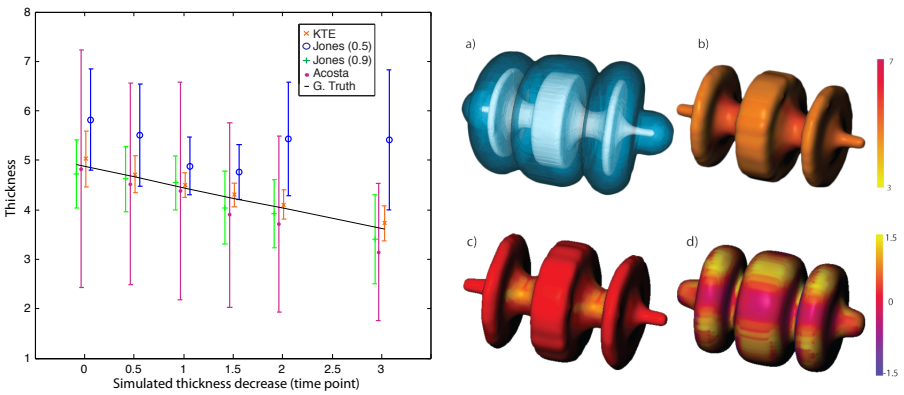


Figure 5.7: Left: Mean and standard deviation of the estimated thickness at time-points 1 to 3 when compared to the ground truth in black. Right: a) The 3D phantom reconstruction, b) the surface with the estimated thickness for time-point 0 with KTE, and c) the difference between the ground truth thickness and the estimated one at time-point 0 for KTE and d) for Acosta et al.

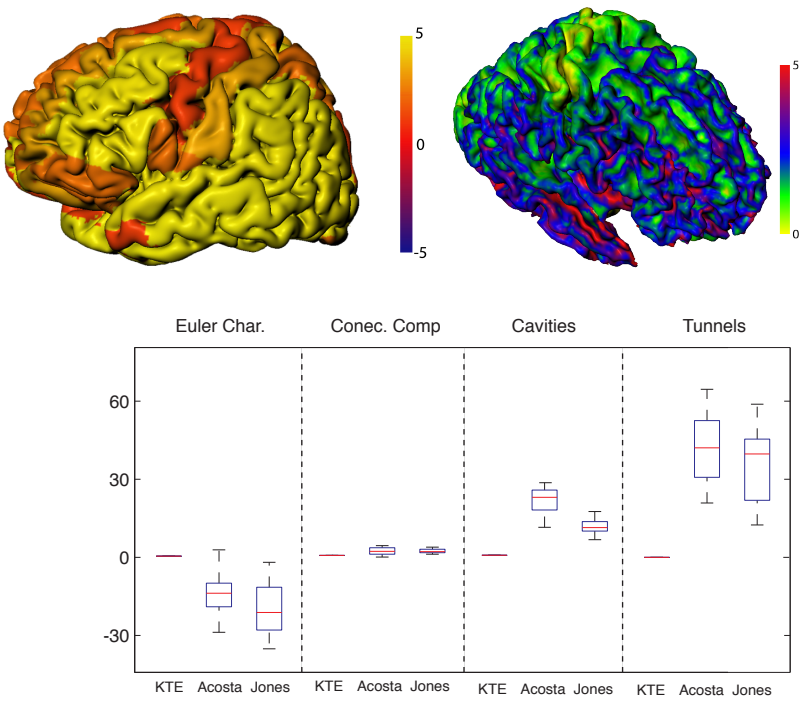


Figure 5.8: Top left: Log of the p-values of statistical difference between AD patients and controls, where positive and negative values represents thinning and thickening of the cortex of AD patients when compared to controls, using KTE; Top right: An example of an unsmoothed topologically correct thickness surface from the AIBL database; Bottom: The Euler characteristic and the number of connected components, cavities and tunnels for the estimated thickness surface

Brain MRI analysis

To further investigate the ability of the proposed method to extract topologically correct measurements from real world segmentations, the thickness of the cortical layer was calculated on the AIBL dataset. The purpose of this study was to evaluate group separation between the cortical thickness of controls and Alzheimer's Disease (AD) diagnosed patients. From the full database, a subset of 54 AD diagnosed patients and 54 age- and gender-matched controls with T1-weighted 3T MRI volumetric images acquired using a 3D MPRAGE sequence (typically $1.20 \times 1.00 \times 1.00mm$) was selected. Images were segmented using an Expectation-Maximization based probabilistic framework of section 3.1 and thickness was calculated with $t = 0.05$ equivalent to 5% probability. The segmentations were then parcellated into different areas using the AAL atlas and two-tailed unequal-variance two-group t-tests were calculated between AD patients and controls over each AAL region.

Fig. 5.8(top left) shows the statistical difference between AD patients and controls in the different areas of the brain projected onto the AAL template, using the proposed method. It shows an overall reduction of the average cortical thickness in AD patients and also an expected symmetrical pattern of cortical lost, with highly statistically significant differences in thickness on the temporal, occipital and mid-frontal areas, clinically known to be affected in AD. More importantly, Fig. 5.8(bottom) shows that the surface with the estimated thickness is homotopic to a sphere for KaTE (Euler characteristic of 2) while both Acosta et al. (2009) and Jones et al. (2000) obtain surfaces with inconsistent topologies.

5.1.4 Conclusions

In this section, a new method to extract measurements of thickness from objects that are homotopic to a sphere is presented. The main contribution of this work lies in three points: Khalimsky Cubic Complex based homotopic operations are used to correct the topology of the segmentation; a multi-stage Laplacian is then used to differentiate between pure from PV voxels; finally, the proposed method achieves sub-voxel accuracy without the use of *ad hoc* ray casting techniques by integrating through PV voxels following the Laplacian streamline field.

Experiments on digital phantoms with known ground truth thickness shows that the proposed method is more accurate and precise than two state of the art methods in retrieving a thickness measurements, while enforcing spherical topology. Quantitative analysis on brain data showed that the proposed algorithm is able to retrieve highly significant group differences between the cortical thickness of AD patients and controls, while maintaining spherical topology.

5.2 4D-KaTE: Longitudinal Cortical Thickness Estimation using Khalimsky's Cubic Complex

Longitudinal measurements of cortical thickness is a current hot topic in medical imaging research. Measuring the thickness of the cortex through time is normally hindered by the presence of noise, PV effects and topological defects, but mainly by the lack of a common directionality in the measurement to ensure consistency. In this section, I propose to extend the previous section to a 4D pipeline (3D + time), again using the Khalimsky cubic complex for the extraction of a topologically correct Laplacian field in an unbiased temporal group-wise space. The thickness at each time point is then obtained by integrating the probabilistic segmentation (transformed to the group-wise space) modulated by the Jacobian determinant of its deformation field through the group-wise Laplacian field. Experiments performed on digital phantoms show that the proposed method improves the time consistency of the thickness measurements with a statistically significant increase in accuracy when compared to two well established 3D techniques and a 3D version of the same method. Furthermore, quantitative analysis on brain MRI data showed that the proposed algorithm is able to retrieve increasingly significant time consistent group differences between the cortical thickness of AD patients and controls.

5.2.1 Derived publications and methodological sumary

- **Cardoso, M. J**., Clarkson, M.J., Modat, M., Ourselin, S.: Longitudinal Cortical Thickness Estimation using Khalimsky's Cubic Complex. (2011) MICCAI

Here, I extend the method described in Cardoso et al. (2011a) and also in section 5.1, in order to encompass 4D consistency. Here, all the time points are registered to an unbiased temporal group-wise space and a group-wise segmentation is then obtained by means of a multivariate EM segmentation algorithm, described in section 3.1, specifically designed for cortical thickness estimation. A multistage Laplacian is then calculated on the temporal group-wise space, and the *per* time point values of thickness are then obtained by integrating over the single time point segmentations along the temporal group-wise derived streamlines.

5.2.2 Method

Proposed Pipeline

The pipeline of most 3D cortical thickness algorithms can be described in 3 steps:

1. The image is segmented into several classes;
2. An implicit or explicit correspondence from one side of the cortex to the opposite side is then found using a multitude of methods;
3. Finally, the thickness is measured according to a specific metric.

Any of these steps can be altered in order to introduce temporal consistency. However, altering steps 1 or 3 will directly affect the measurement of thickness, as either the segmentation or the thickness metric itself would be affected by a constraint. This can reduce the statistical significance of difference

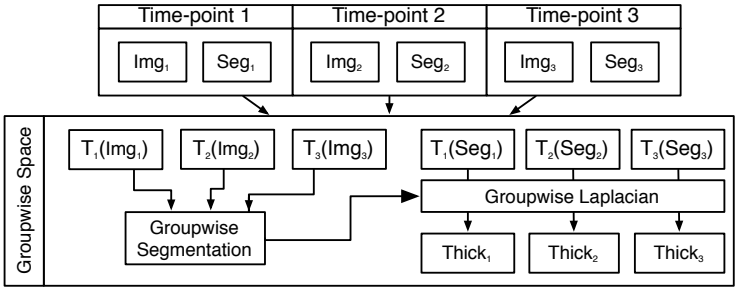


Figure 5.9: Top) Individual time points are segmented and registered to the temporal group-wise space; Bottom) A group-wise segmentation is obtained from the transformed images and used to create a group-wise Laplacian field. Thickness is obtained by integrating through the group-wise Laplacian field over the transformed segmentation at each time point.

measurements between groups, leading to increased sample sizes. In order to maintain the sensitivity of the measurement intact but still achieve temporal consistency, we propose to constrain only the direction of the measurement, by changing step 2. The proposed 4D cortical thickness pipeline consists of five steps, as shown on Fig.5.9. For the sake of simplicity, I assume a series of skull stripped brain images acquired at 3 different time points. The five steps can the be described as:

1. The images are segmented independently using a previously published voxel based probabilistic segmentation algorithm (Cardoso et al., 2011b) specifically designed for cortical thickness. This segmentation will separate the brain into 5 classes: WM, cortical and deep GM and internal and external CSF. The cerebellum is removed within the same pipeline by atlas propagation. Here, the segmentations from the deep GM and internal CSF are added to the WM segmentation in order to create a class containing all the internal structures, simply called WM for the sake of clarity.

2. An unbiased group-wise registration is then created between all the time points. This iterative registration (Rohlfing et al., 2004a) process results in a transformation $\mathbf{T}_t$ for each time point t to the average group-wise space. All the skull stripped images and respective segmentations are transformed to this space.

3. In order to create a group-wise segmentation, a multivariate version of the same algorithm (Cardoso et al., 2011b) is used. Here, the segmentation model assumes that each label is not only a realisation of one image (time point) but a combined multivariate realisation of all time points, leading to a segmentation with a high level of cortical detail.

4. A topologically correct Laplacian field map is created using the group-wise segmentation as described in section 5.2.2. This Laplacian map has the directionality information derived from the group-wise segmentation.

5. This common directionality is then used at each time point to drive a PDE based thickness measure with a speed function proportional to the transformed segmentation modulated by the Jacobian of this transformation, as described in section 5.2.2

The topologically correct Laplacian field in step 4 constraints the time consistent direction of thickness measurement. This consistency is important as the correspondences between both sides of the cortex can change dramatically on simple 3D models due to sulci and gyri opening and closing.

Multi-stage Laplace equation on the group-wise space

In order obtain a 4D consistent measurement of thickness at each time point, a unique association between two sides of the cortex is required. For this purpose, I use the Laplace equation (as in section 5.1), a second order partial differential equation (PDE), solved between two enclosed boundaries Ω and Ω' in the group-wise space. Similarly to section 5.1 and (Cardoso et al., 2011a), instead of a single Laplacian field for all the pure voxels as in (Jones et al., 2000; Yezzi Jr and Prince, 2003; Acosta et al., 2009), a multiple Laplacian field is solved. This obviates the problems regarding the estimation of surface normals for PV integration using ray casting (Acosta et al., 2009). A set of isolines is generated for each tissue type from the topologically correct group-wise segmentation. The Laplace equation is then solved between these four equipotential lines resulting in a smooth transition field traversing the cortex. The solution of this Laplace equation under an anisotropic voxel grid in the Khalimsky space is presented in (Cardoso et al., 2011a). From the resultant Laplacian field, the normals to the direction of the Laplacian isolines in the group-wise space, denoted by N^{GW}, are calculated using finite differences.

Thickness measurement at each time point

In order to measure thickness, the length of the streamlines between the inner and outer surface has to be measured at each time point by integrating the vector field N^{GW} on the group-wise space. Because digital topology is not preserved even under diffeomorphic transformations, each time point has to be deformed to the group-wise space using the previously computed group-wise transformation. In order to measure thickness, instead of the basic form partial differential equation proposed by Yezzi Jr and Prince (2003), where the speed of the advancing front is assumed to be 1, we use a more generalised form of the PDE. Here $\nabla L^t \cdot N^{\text{GW}} = f^t$, for an unknown function L^t at time-point t and assuming that N^{GW} and f^t are known. As stated before, and differently from (Jones et al., 2000; Yezzi Jr and Prince, 2003; Rocha et al., 2005; Acosta et al., 2009), the value of f will be spatially varying and equal to the probability of belonging to the cortical GM modulated by the Jacobian determinant of the transformation, $f^t = p^t_{\text{GM}}|\mathbf{T}(x)^t|$. Note that one should only use the Jacobian component in the direction of T. However, as the transformation between different time points is mostly in the direction of T, the use of the Jacobian determinant is a good approximation. This value will act as *time cost* and will make the value of L^t equivalent to the time of arrival in a level-set framework. Modulation by the Jacobian determinant is necessary in order to take the voxel compression into account. Even though collisions of the advancing front might exist, they are not a problem due to the upwind nature of the integration and the existence of the group-wise vector field N^{GW}. Let $L^t_{0_{(x,y,z)}}$ be a function that measures the time of arrival (arc length of the streamline according to the time cost f^t) from the boundary of set $F_{WM_{\text{pure}}}$ on the group-wise space to a point in the object, and $L^t_{1_{(x,y,z)}}$ be the time of arrival from the boundary of set $F_{\text{CSF}_{\text{pure}}}$, again in the group-wise space, to the point in the object. The values of L_0 and L_1 are calculated using anisotropic finite differences, as described in section 5.1, and the final value of thickness is then

defined as $Thick^t = L_0^t + L_1^t$. In order to reduce the bias of any further statistical analysis, the value of thickness is only calculated at mid-isopotential line on the group-wise Laplacian field.

5.2.3 Experiments and Results

This experimental section is divided into two sub-sections. First, a digital phantom with time evolving ground truth thickness is used to assess the accuracy and sensitivity of the proposed algorithm compared to three 3D state-of-the-art methods. The proposed method is then applied to brain MRI data in order to assess group separation in terms of cortical thickness between Alzheimer's disease diagnosed patients and controls.

Phantom validation

In order to evaluate if the proposed method can accurately retrieve the underlying thickness of an object, 7 folded 3D digital phantoms with spherical topology and known ground truth thickness were created (Fig. 5.10), resulting in six high resolution isotropic images with 3 structures equivalent to WM, GM and CSF. The thickness of the object is changing with time from 5.2 to $3.6mm$ in order to simulate a thickness loss in an object over several time points. Note that the sulci will open after time-point 1. These high resolution phantoms were then down-sampled by 5 in order to simulate PV effect and the thickness of the down-sampled structures was then measured. I compare the proposed 4D method with the 3D version of the same algorithm, the method proposed by Jones et al. (2000) and the method proposed by Acosta et al. (2009).

Results show that all 3D methods are highly sensitive to temporal structural changes. When compared to the ground truth, the thickness change in time is overestimated, possibly due to sulci opening. The proposed method, on the other hand, uses a 4D consistent directionality derived from all the time-points, resulting in a much more accurate and precise thickness estimation. One should be cautious when reporting cortical thickness loss in time, as 3D methods can severely overestimate it.

Brain MRI analysis

To further investigate the temporal consistent of the proposed method, the thickness of the cortical layer was calculated on the ADNI dataset at 3 time points. The main purpose of this study was not to evaluate group separation between different groups but to assess their stability in time. From the full ADNI

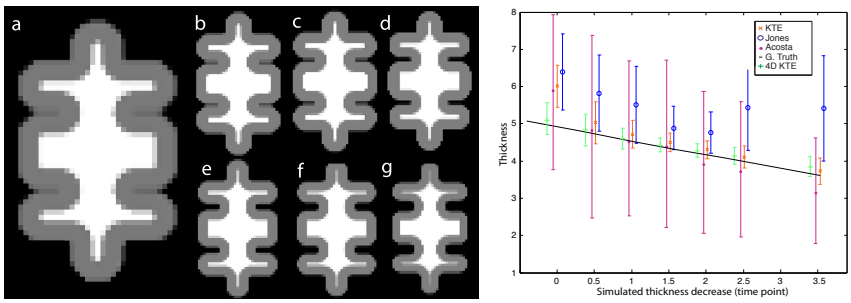


Figure 5.10: Left: A 4D simulated high resolution phantom with a time varying thicknesses ranging from $5.2mm$ to $3.6mm$ (a-g). Right: Mean and standard deviation of the estimated thickness at all voxel positions from timepoint 1 to 3.5 when compared to the ground truth in black.

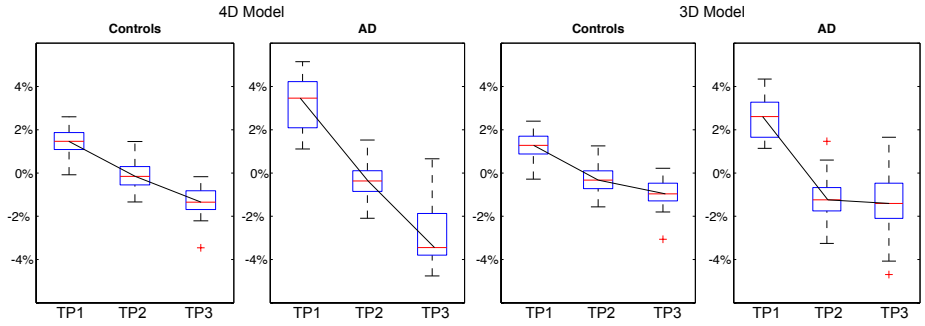


Figure 5.11: A plot showing the progression of the average cortical thickness in time, normalised to the average thickness over all time points. From left to right: The normalised thickness for controls and AD patients using the 4D (proposed) method and 3D (time independent) versions of the algorithm.

dataset, a subset of 60 age- and gender-matched subjects (30 AD and 30 controls) were selected. Each subject has T1-weighted 1.5T MRI volumetric images acquired using a 3D MPRAGE sequence (typically $1.20 \times 1.00 \times 1.00mm$) at 0, 12 and 24 months. Fig. 5.11 shows the progression of the distribution of the average value of thickness within the cortex at each time point, normalised to the average thickness over all time points. Due to the lack of time consistency, unexpected inversions of the thinning pattern occur on the 3D version of the algorithm. The 4D version of the algorithm shows a marked improvement regarding the stability of the thinning pattern. This leads to a reduction of the standard deviation of the thickness distributions within each group, increasing the statistical power. In order to compare the different groups (AD and controls) on a per area basis, the group-wise space was parcelated into different areas using the an anatomical atlas. The 3D method shows statistically significant differences in thickness ($p < 10^{-3}$) on both the temporal and parietal region. The frontal region is statistically significantly thinner ($p < 10^{-3}$) at both time-points 1 and 3 but it but does not achieve the significance threshold at time-point 2. On the other hand, the proposed 4D method shows statistically significant differences in thickness at the level $p < 10^{-5}$ in the middle and inferior temporal and parietal regions and $p < 10^{-3}$ in the frontal gyrus region in the first time point. From time point 2 onwards, the frontal region becomes significant at $p < 10^{-5}$ and both the superior and occipital regions become statistically significant at $p < 10^{-3}$.

Due to the lack of 4D consistency in the segmentation, cortical lost might be over-estimated. More work still has to be done to future explore the use of a 4D segmentation step for improved consistency. One should also consider the use of the full Jacobian matrix for the modulation step instead of it's determinant.

5.2.4 Conclusions

In this section, I present a new method to extract 4D measurements of thickness from cortical segmentations. First, all the time points are registered to an unbiased and temporal consistent group-wise space. Then, a time consistent group-wise point-to-point correspondence is found by means of a multistage Laplacian field derived from a multivariate segmentation in the group-wise space. This common directionality is then used to calculate the thickness at each time point.

Experiments on digital phantoms with known ground truth thickness show that the proposed method is more accurate and precise at retrieving true thickness values than other previously published methods, thereby reducing the over-estimation of cortical thinning in the presence of sulci opening. Quantitative analysis on brain MRI data showed that the proposed algorithm is able to retrieve increasingly significant time consistent group differences between the cortical thickness of AD patients and controls.

Chapter 6

Open Software Effort

The open source movement, as exemplified by the culture that has evolved around the Unix/Linux based operating systems, hypothesises that open, non-commercial sharing and co-development of computer code and algorithms, leads to better and less error prone technical solutions than the prevailing closed-source/copyrighted/business paradigm. Open source is a development approach that promotes transparency and promises more quality, reliability and flexibility in the production and testing of software. Due to their open nature, open source licenses normally allow anyone to contribute, understand, refactor and reuse the code with no restrictions. Many arguments have been made throughout the years regarding the benefits and problems of this new approach. A recent exposure in Nature by Ince et al. (2012) proposed a series of advantages on the context of open-source/open-science:

- Software can be inconsistent and can contain errors. Over the years, researchers have quantified the occurrence rate of such defects to be approximately one to ten errors per thousand lines of source code.

- Describing algorithms using mathematical or natural language can be limiting. Formal mathematical description commonly focuses on the model and not on the implementation details, while natural language can have ambiguous algorithmic interpretations.

- Without open source code, direct reproducibility of the results is impossible. Time-consuming reprogramming of published algorithms and model is most of the times the only way to check the validity of the results presented by the paper.

- Without a detailed description of the full algorithmic assumptions, one cannot check for statistical validity and explain any discrepancies between implementations.

While the advantages presented by Ince et al. (2012) relate to the correctness of the methods and their scientific reproducibility, several other aspects can be beneficial. First, open-source software enables researchers to build upon other research. Second, released code provides excellent teaching material, both to the post-graduate and undergraduate levels, as they are realistic source code examples. Third, it encourages non-institutional researchers to get involved in research (citizen science). For this to happen both code and data transparency is needed. Last but not least, open-souce code enables a syn-

ergic interaction between the original developer and the community, resulting in improvements in both functionality, accuracy and robustness.

As a supporter of this approach, all the code developed during my PhD is available under the new Berkeley Software Distribution (BSD) license.

With a BSD license, redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the copyright notice, the license's list of conditions and the license's disclaimer.
- Redistributions in binary form must reproduce the copyright notice, this license's list of conditions and the license's disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the organisation nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

6.1 NiftySeg

The software project, called NiftySeg, part of the NifTK suite of software (`http://niftyseg.sf.com`). This project contains programs to perform EM based segmentation of NIFTI or ANALYSE images. It also contains a package of label fusion algorithms (MV, STAPLE, SBA) with different types of ranking strategies. The software for cortical thickness estimation is not available online at the time of submission as further work is necessary to make it easy to use.

To download the latest version, please check out the code by copying the following line to the terminal:

```
svn co https://niftyseg.svn.sourceforge.net/svnroot/niftyseg niftyseg
```

or from `http://sourceforge.net/projects/niftyseg/`

It currently features 3 main functions that are part of the methodological contribution of this thesis:

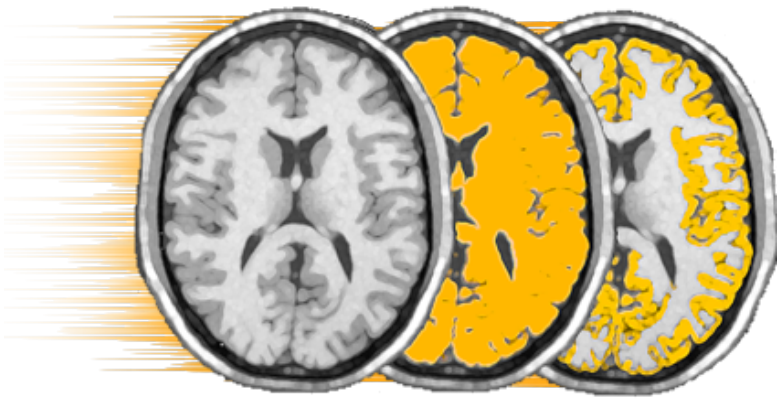


Figure 6.1: The logo of the NiftySeg software, available at `http://niftyseg.sf.com`.

- seg_LoAd: the LoAd segmentation algorithm, presented in Section 3.1
- seg_EM: a general purpose and highly tweakable EM segmentation algorithm, containing the code specific for Section 3.2
- seg_LabFusion: single and multi-label fusion package, containing the code used for Chapter 4

There are also two general purpose functions:

- seg_maths: a image operation tool that allows simple arithmetic and morphological operation to be performed sequentially on images.
- seg_stats: a simple image statistic calculation tool (classical statistics, ranges, overlaps, ROI localisation, etc.)

Underneath, there are 3 main libraries _seg_LoAd, _seg_EM and _seg_LabFusion making up the main application programming interface (API) for external linking. Each one of these libraries defines a C++ object that defines, sets and runs the model and the necessary variables. A few other lower level functions are also available, but their usage requires much deeper code integration.

6.1.1 seg_LoAd

This executable presents the standard implementation of the methods presented in Section 3.1. As stated before, it describes an EM algorithm with INU correction, an MRF for spatial smoothness with spatially-adaptive strength, an explicit PV model and a prior relaxation section that reduces the bias introduced by the population atlas.

The validation presented in Section 3.1 used this executable, will all the parameters set to default. Thus, the command line being used was

```
seg_Load -in T1.nii -priors WM.nii GM.nii CSF.nii dGM.nii iCSF.nii
         -out OutputSegmentation.nii}
```

Usage message

```
Usage: seg_Load -in <filename> [OPTIONS].

* * Mandatory * *
-in <filename>       Filename of the input image image
-mask <filename>     Filename of the brainmask of the input image
-out <filename>      Filename of for the segmented images
* * Options * *
-priors <filenames>  The 5 priors in this order: WM,GM,CSF,dGM,iCSF
-max_iter <int>      Maximum number of iterations (default = 100)
-rf <int>            Relaxation factor 0<RF<1 (default = 1)
-v <int>             Verbose level [0 = off, 1 = verbose, 2 = debug] (0)
-mrf_beta <float>    MRF prior strength [off = 0] (default = 0.1)
-bc_order <int>      Polinomial order for the bias field [off=0,max=6] (5)
-pv_off              Do not perform the PV modeling
-sg_off <int>        Do not improve sulci and gyri deliniation
```

6.1.2 seg_EM

This executable presents a general purpose EM segmentation with algorithmic options like INU correction (-bf_order), an MRF for spatial smoothness (-mrf_beta), outlier detection (-outlier) as described in Van Leemput et al. (2001), anatomical prior relaxation (-rf) and semi-conjugate prior over the model parameters (-MAP) as described in Section 3.2, ROI selection through masking (-mask), all with (-priors) or without (-nopriors) using population priors.

The validation presented in Section 3.2 used this executable, will all the parameters set as described. The command line being used was similar to

```
seg_EM -in T1.nii -priors 6 WM.nii GM.nii CSF.nii dGM.nii
       Pons.nii Cerebelum.nii -rf 0.5 2.0 -MAP <average,
       std of the cluster means> -out Segmentation.nii
```

Usage message

```
Usage: seg_EM -in <filename> [OPTIONS]

* * Mandatory * *
-in <filename>          Filename of the input image
-out <filename>         Filename of the segmented image
                        The input image should be 2D, 3D or 4D images.
                        2D images should be on the XY plane.
                        4D images are segmented as if they were multimodal.

* * Select one of the following (mutually exclusive) * *
-priors <n> <fnames>    The number of priors (n>0) and their filenames.
-priors4D <fname>       4D image with the piors stacked in the 4th dimension.
-nopriors <n>           The number of classes (n>0)

* * General Options * *
-mask <filename>        Filename of the brain-mask of the input image
-max_iter <int>         Maximum number of iterations (default = 100)
-v <int>                Verbose level [0 = off, 1 = on, 2 = debug] (default = 0)
-mrf_beta <float>       MRF prior strength [off = 0, max = 1] (default = 0.4)
-bc_order <int>         Polynomial bias field order [off=0,max=5] (def=3)
-bc_thresh <float>      Bias field correction only if the ratio < bc_thresh
-bc_out <filename>      Output the bias corrected image
-reg <filename>         Amount of regularisation on the covariance
-outlier <fl1> <fl2>    Outlier detection as in (Van Leemput TMI 2003).
                        <fl1> is the Mahalanobis threshold [3 - 7]
                        <fl2> is a convergence ratio below wich the outlier
                        detection is going to be done [recommended 0.001].
-out_outlier <filename> Output outlierness image
-rf <rel> <gstd>        Relax Priors [relaxation factor: 0<rf<1 (recom=0.5),
                        Gaussian regularization: gstd>0 (recom=2.0)] /only 3D/
-MAP <M V M V ...>      MAP formulation: M and V are the parameters
                        of the semiconjugate prior over the class mean
```

6.1.3 seg_LabFusion

A label fusion algorithm presenting the techniques described in Chapter 4 and all the other techniques used to validate the proposed method. It presents 4 fusion strategies (-STEPS, -STAPLE, -MV or -SBA), all of them using either a global (-GNCC), ROI-based (-ROINCC), local (-LNCC) or no image similarity (-ALL). A few other options pertaining the use of an MRF (-MRF_beta), the initial sensitivity and specificity estimates and the use of only non-consensus voxels (-unc) for the STAPLE and STEPS algorithm. All processing can be masked (-mask), greatly reducing memory consumption.

Both validation section in Chapter 4 used this executable, will all the parameters set as described. As an example, the command line being used for STEPS was similar to

```
seg_LabFusion -in 4D_Propragated_Labels_to_fuse.nii
              -out FusedSegmentation.nii -STEPS 2 15
              TargetImage.nii 4D_Propagated_Intensities.nii
```

Usage message

```
Usage: seg_LabFusion -in <filename> -<Type of Label Fusion> [OPTIONS]

* * Mandatory * *
-in <filename>                Filename of the 4D integer label image

* * Type of Classifier Fusion (mutually exclusive) * *
-STEPS <k> <n> <i> <t>        STEPS algorithm
                              Size of the kernel (k), number of local labels to use (n),
                              Original image to segment (3D Image), registered templates (4D Image).
-STAPLE                       STAPLE algorithm
-MV                           Majority Vote algorithm
-SBA                          Shape Based Averaging algorithm (Beta)

* * General Options * *
-v <int>                      Verbose level [0 = off, 1 = on, 2 = debug] (default = 0)
-unc                          Only consider non-consensus voxels to calculate statistics
-out <filename>               Filename of the integer segmented image (default=LabFusion.nii.gz)
-mask <filename>              Filename of the ROI for label fusion (greatly reduces memory usage)
* * STAPLE and STEPS options * *
-prop <proportion>            Proportion of the classifier (automatically estimated by default)
-prop_update                  Update label proportions at each iteration.
-setPQ <P> <Q>                Value of P and Q [ 0 < (P,Q) < 1 ] (default = 0.99 0.99)
-MRF_beta <float>             MRF prior strength [ 0 < beta < 5 ]
-max_iter <int>               Maximum number of iterations (default = 50)
-conv <float>                 Ratio for convergence (default epsilon = 10^-6)

* * Ranking for STAPLE and MV (mutually exclusive) * *
-ALL  (default)               Use all labels with no Ranking
-GNCC <n> <img> <tmpl>        Global Normalized Cross Correlation Ranking:
                              Number of sorted classifiers to use (n),
                              Original image to segment (3D image), registered templates (4D Image).
-ROINCC <d> <n> <img> <tmpl>  ROI Normalized Cross Correlation Ranking:
                              Dilation of the ROI ( <int> d>=1 ), Num classifiers to use (n),
                              Original image to segment (3D image), registered templates (4D Image).
-LNCC <k> <n> <img> <tmpl>    Locally Normalized Cross Correlation Ranking):
                              Size of the kernel (k), number of local classifiers to use (n),
                              Original image to segment (3D Image), registered templates (4D Image).
```

6.1.4 seg_maths

The executable seg_maths enables the sequential execution of arithmetic operations, like multiplication (-mul), division (-div) or addition (-add), binarisation (-bin) or thresholding (-thr) operations and convolution by a Gaussian kernel (-smo). It also alows mathematical morphology based operations like dilation (-dil), erosion (-ero), connected components (-lconcomp) and hole filling (-fill), Euclidean (-euc) and geodesic (-geo) distance transforms, local image similarity metric calculation (-lncc and -lssd). Finally, it allows multiple operations over the dimensionality of the image, from merging 3D images together as a 4D image (-merge) or splitting (-split or -tp) 4D images into several 3D images, to estimating the maximum, minimum and average over all time-points.

Usage message

```
Usage: seg_maths <input> <operation> <output>.

* * Operations on 3-D and 4-D images* *
-mul   <float/file>  Multiply image <float> value or by other image.
-div   <float/file>  Divide image by <float> or by other image.
-add   <float/file>  Add image by <float> or by other image.
-sub   <float/file>  Subtract image by <float> or by other image.
-pow   <float>       Image to the power of <float>.
-thr   <float>       Threshold the image below <float>.
-uthr  <float>       Threshold image above <float>.
-smo   <float>       Gaussian smoothing by std <float> (up to 4-D).
-sqrt                Square root of the image.
-exp                 Exponential root of the image.
-recip               Reciprocal (1/I) of the image.
-abs                 Absolute value of the image.
-bin                 Binarise the image.

* * Operations on 3-D images * *
-dil   <int>         Dilate the image <int> times (in voxels).
-ero   <int>         Erode the image <int> times (in voxels).
-lconcomp            Take the largest connected component
-fill                Fill holes in binary object
-euc                 Euclidean distance trasnform
-geo <float/file>    Geodesic distance given speed function <float/file>

* * Dimensionality reduction operations: from 4-D to 3-D * *
-tp <int>            Extract time point <int>
-tpmax               Get the time point with the highest value
-tmean               Mean value of all time points.
-tmax                Max value of all time points.
-tmin                Mean value of all time points.

* * Dimensionality increase operations: from 3-D to 4-D * *
-merge <i> <d> <f> Merge <i> images and current img in <d>-dimension
-splitlab            Split the integer labels into multiple timepoints

* * Image similarity: Local metrics * *
-lncc <file> <std>   Local CC between img and <file> on a kernel with <std>
-lssd <file> <std>   Local SSD between img and <file> on a kernel with <std>

* * Output * *
-odt <datatype>      Set output <datatype> (<u>char, <u>short, <u>int, float, double).
```

6.1.5 seg_stats

The executable seg_stats allows the estimation of several basic image statistics, like the average (-a), std (-s), percentiles (-p), the volume of probalisistic (-V) or binary (-v) segmentations, geometrical localisation of objects, like the bounding box (-B), centre of mass (-c) and even the Dice overlap (-d) for one or multiple labels. All these metrics can be calculated on the full image, on a restrained subset defined as a binary mask (-m) or only for locations in the image above a certain intensity value (-t)

Usage message

```
Usage: seg_stats <in> [constraints] [statistics]

* * Constrains (optional) * *
-m <mask>       Only estimate statistics within the masked area.
-t <float>      Only estimate statistics if voxel is larger than <float>.

* * Range operations (datatype: all) * *
-r              Range <min max> of all voxels.
-R              Robust range (assuming 2\% outliers) of all voxels
-p <float>     <float>th percentile of all voxels intensity [0,100]

* * Classical operations (datatype: all) * *
-a              Average of all voxels
-s              Standard deviation of all voxels
-v              Volume of all binarized voxels (<# voxels> * <vox. vol>)
-V              Volume of all probabilsitic voxels (sum(<in>) * <vox. vol>)
-n              Sum of all binarized voxels (<# voxels>)
-N              Sum of all probabilsitic voxels (sum(<in>))

* * Coordinates operations (datatype: all) * *
-x              Location (in vox) of the smallest value in the image
-X              Location (in vox) of the largest value in the image
-c              Location (in vox) of the centre of mass of the object
-B              Bounding box of all nonzero voxels

* * Label attribute operations (datatype: char or uchar) * *
-d <in2>        Dice score between all classes in <in> and <in2>
-D <in2> <csv>  Same as above but save results in CSV file.
```

Chapter 7

Conclusion

This thesis presents a full suite of methods and pre-processing steps required to obtain robust, accurate and longitudinally stable measurements of structural volume and cortical thickness in normal, pathological and morphologically challenging subjects. Nonetheless, most of the tools developed in this thesis, mainly the ones on tissue segmentation and brain parcellation, have a much broader spectrum of applications than cortical morphology characterisation.

In Chapter 3, the problem of robust, accurate and unbiased brain tissue segmentation of elderly subjects was addressed with a probabilistic segmentation framework (named LoAd) optimised using an Expectation-Maximisation algorithm. This framework introduces three novel modifications to the probabilistic framework in order to: reduce the influence of the population priors in an anatomically coherent way, improve the PV estimation, and better delineate the deep sulci and gyri. These small modifications in the segmentation pipeline are shown to have a significant impact on, not only on the accuracy and robustness of the segmentation, but mainly on the estimation of cortical thickness.

The knowledge acquired during the development of the LoAd segmentation technique has enabled the adaptation of the framework to preterm neonatal brains. Due to the presence of severe pathologies in preterm neonatal brains (e.g. ventriculomegaly, cysts, sub-arachnoid CSF, etc...), a new atlas relaxation strategy was used in order to adapt the normal population priors to pathological cases. As this atlas relaxation strategy introduced too much flexibility, the model had to be constrained using a semi-conjugate prior over the parameters. The combination of both strategies provided enough flexibility to adapt the population atlas to pathological brains while ensuring the correct convergence of the segmentation. This methodology has allowed, for the first time, the automated morphometric analysis of severely pathological preterm brains and is currently being used to explore the structure of the preterm cortex and test several hypothesis regarding the correlation between structural abnormalities and delays or disruptions in normal cortical maturation.

Chapter 4 proposed two solutions for the problem of skull stripping and parcellation of multiple brain structures as a pre-processing step for tissue segmentation, morphometric studies and region-based cortical thickness estimation. A probabilistic framework for label fusion was first explored in order to parcelate the brain into multiple simultaneous structures. This algorithm extends the classic STAPLE framework, making the label fusion locally selective, spatially smooth and unbiased to the size of the

ROI. This technique provides very high accuracy for the segmentation of key brain structures (mainly the hippocampus) and is currently being used as part of the Dementia Research Centre processing pipeline for image analysis and clinical trials.

As the performance of this algorithm is mainly limited by the quality and morphological variability of the samples in the template database, a new algorithm (GIF: Geodesic Information Flows) that can parcelate pathological and morphologically off-database subjects was also proposed. This framework explored the concept of graph information parsing using an implicit local manifold structure where highly pathological subjects are parcelated by first propagating the parcellations to locally intermediate morphologies. This procedure extends and generalises the idea of label fusion to the much broader concept of information diffusion, a concept that allows the propagation of floating point data, probabilistic data, integer labels and even images. Thus, the GIF framework allows the generation of patient specific probabilistic tissue atlases, brain masks, structural parcellations, bias-field correction, and much more, in a single unified framework.

Finally, Chapter 5 proposed the use of a new framework, in conjunction with the previously described tools, for estimating cortical thickness with spherical topology assumptions. This new framework (named KaTE), based on the Khalimsky cubic complex, estimates cortical thickness by calculating the integral over a cortical density field of a Laplacian streamline that traverses the cortex This framework not only improves the accuracy and precision of the cortical thickness estimation, but also corrects the topology of the cortical segmentation by using the topologically-invariant collapse operation on the cubic complex set. The KaTE algorithm achieved a significant reduction in cortical thickness estimation error and measurement dispersion, thus increasing the power to separate groups (i.e. controls and AD patients).

This technique was then extended to longitudinal studies though the use of a time-wise minimum deformation space combined with a Jacobian modulation step on the density field. This longitudinal extension provides time-consistent cortical thickness estimates without constraining the measurements themselves, thus resulting in an increase in statistical power for population studies with no constrain bias.

Overall, this thesis explores several crucial pre- and post-processing steps for neuroimage analysis, enabling the study of cross-sectional and longitudinal structural morphometry and cortical thickness. The work developed in this thesis has demonstrated that carefully built algorithms can result in a good accuracy and precision with minimal bias.

7.1 Future Research Directions

Notwithstanding the improvements shown in this thesis, there are still many derived problems and associated conceptual ideas that are worth pursuing, either by myself or anyone in the medical imaging community. From all these, I will describe two impactful ideas that can change the way we currently analyse medical images.

My main worry at the time of conclusion of this thesis regards the inherent bias of doing groupwise analysis and relying on groupwise spaces. Groupwise spaces, normally defined as minimum morpholog-

ical/deformation spaces, were the natural extension to the use of a single subject as the chosen space for image comparison due to their feasible computational complexity and error optimality. However, even though groupwise spaces should minimise the mapping errors for image comparison, they still suffer from two main problems. First, although the mapping errors are in theory minimised in a groupwise space, no studies have shown that they are actually close to zero. Thus, structural mismatch due to local morphological variability can, and most probably will, introduce spurious findings. If the findings are not caused by registration errors, then they should be independent of the actual space used for comparison. Thus, is it possible to devise a strategy where all the images are mapped to every image and then compared? Is the idea that "a true finding has to be independent of the space used for comparison" correct? The second point, which falls into the same argument, is the idea that all data should be compared in an equal manner. For example, if two different groups of control subjects that belong to two different morphological subgroups (e.g. with different sulcal patterns) are compared, significant differences in grey matter density and cortical thickness will be found between them even though both groups are cognitively normal. The same effect can occur between a normal and a pathological population due to small sample sizes or different distributions of non-pathological morphologies. Thus, the result will show both the normally occurring and pathologically derived differences between groups. In theory, both these differences can be decoupled by stratifying data according to their non-pathological morphological characteristics. The data can then be locally compared within each morphological subgroup. This stratification procedure will also ensure that the registration can cope with the mapping between images that are being compared and that the mapping error is actually close to zero rather than only minimised. However, as the belongingness to specific morphological subgroups is not a binary assignment but rather a continuum, is it feasible to compare only between the most morphologically similar subjects without explicitly grouping them and without loosing all the statistical power? Is it possible to decouple the natural morphological differences between normal subjects and pathological ones? All these questions are still very open points of discussion and are many times ignored due to availability cascade bias ("If this is what everyone is doing, then it should be correct"), confirmation bias (evaluating evidence that supports one's preconceptions differently from evidence that challenges these convictions) and bias due to hypothesis-generating research (the formulation of hypothesis that support the measurements).

My second main methodological worry and possible aim of future work pertains with the current lack of connection between macro-structural morphology and tissue microstructure. As the size, shape and thickness of brain structures are only surrogate biomarkers of the underlying biology and pathology and not true direct biomarkers, one would ideally estimate the properties of the tissue at the microstructural level. Recently, several authors have shown that diffusion weighted imaging can provide insightful information about tissue micro-structural properties like neuronal density, axon dispersion and extra-/intra-cellular volume ratio. Thus, one is now able to combine and explore the relationship between macro- and micro-structure, further increasing the understanding of pathological changes. Is it beneficial to estimate cellular density and mass over a specific ROI rather than estimating it's volume? Is cognitive performance more correlated with the size of the structure or with cell density? Is a thick

cortex with low neuronal density healthy? Are patients with high neuronally dense cortices more or less at risk of cognitive decline? Is one able to increase the statistical power and reduce the intrinsic ”natural variability” of morphometric studies (e.g. VBM, TBM, atrophy and cortical thickness estimates) by probing the real underlying pathology?

Overall, these two methodological concerns match my current conviction that one needs to decouple pathological and non-pathological micro- and macrostructural changes in order to understand the true underlying biological process. Thus, I believe that Multi-X (a term used by Frederik Maes to describe the use of multi-scale, multi-modality, multi-feature) image understanding will become not only the *de facto* standard but the only clinically acceptable way for disease characterisation.

References

Acosta, O., Bourgeat, P., Fripp, J., Bonner, E., Ourselin, S., Salvado, O., 2008. Automatic Delineation of Sulci and Improved Partial Volume Classification for Accurate 3D Voxel-Based Cortical Thickness Estimation from MR. In: Medical Image Computing and Computer-Assisted Intervention. pp. 253–261.

Acosta, O., Bourgeat, P., Zuluaga, M. A., Fripp, J., Salvado, O., Ourselin, S., Alzheimer's Disease Neuroimaging Initiative, Oct. 2009. Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. Medical Image Analysis 13 (5), 730–743.

Aganj, I., Sapiro, G., Parikshak, N., Madsen, S. K., Thompson, P. M., 2008. Segmentation-free measurement of cortical thickness from MRI. In: IEEE International Symposium on Biomedical Imaging. pp. 1625–1628.

Ajayi-Obe, M., Saeed, N., Cowan, F. M., Rutherford, M. A., Edwards, A. D., Sep. 2000. Reduced development of cerebral cortex in extremely preterm infants. Lancet 356 (9236), 1162–1163.

Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., Rueckert, D., Jul. 2009. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46 (3), 726–738.

Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C., Aug. 2009. Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Transactions on Medical Imaging 28 (8), 1266–1277.

Ashburner, J., Friston, K. J., 2005. Unified segmentation. NeuroImage 26 (3), 839–851.

Ashburner, J., Friston, K. J., 2009. Computing average shaped tissue probability templates. NeuroImage 45 (2), 333–341.

Aubert-Broche, B., Griffin, M., Pike, G. B., Evans, A. C., Collins, D. L., Nov. 2006. Twenty new digital brain phantoms for creation of validation image data bases. IEEE Transactions on Medical Imaging 25 (11), 1410–1416.

Barnes, J., Foster, J., Boyes, R. G., Pepple, T., Moore, E. K., Schott, J. M., Frost, C., Scahill, R. I., Fox, N. C., May 2008. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. NeuroImage 40 (4), 1655–1671.

Battin, M. R., Maalouf, E. F., Counsell, S. J., Herlihy, A. H., Rutherford, M. A., Azzopardi, D., Edwards, A. D., Jun. 1998. Magnetic resonance imaging of the brain in very preterm infants: visualization of the germinal matrix, early myelination, and cortical folding. Pediatrics 101 (6), 957–962.

Boardman, J. P., Craven, C., Valappil, S., Counsell, S. J., Dyet, L. E., Rueckert, D., Aljabar, P., Rutherford, M. A., Chew, A. T. M., Allsop, J. M., Cowan, F. M., Edwards, A. D., 2010. A common neonatal image phenotype predicts adverse neurodevelopmental outcome in children born preterm. NeuroImage 52 (2), 409–414.

Cachier, P., Bardinet, E., Dormont, D., Pennec, X., Ayache, N., Mar. 2003. Iconic feature based nonrigid registration: the PASHA algorithm. Computer Vision and Image Understanding 89 (2-3), 272–298.

Cardoso, M. J., Clarkson, M. J., Modat, M., Ourselin, S., 2011a. On the Extraction of Topologically Correct Thickness Measurements using Khalimsky's Cubic Complex. In: Information Processing for Medical Imaging. pp. 159–170.

Cardoso, M. J., Clarkson, M. J., Modat, M., Ridgway, G. R., Ourselin, S., 2010. Locally weighted Markov random fields for cortical segmentation. In: IEEE International Symposium on Biomedical Imaging. pp. 956–959.

Cardoso, M. J., Clarkson, M. J., Ridgway, G. R., Modat, M., Fox, N. C., Ourselin, S., The Alzheimer's Disease Neuroimaging Initiative, Jun. 2011b. LoAd: A locally adaptive cortical segmentation algorithm. NeuroImage 56 (3), 1386–1397.

Cardoso, M. J., Leung, K. K., Modat, M., Barnes, J., Ourselin, S., Sep. 2011c. Locally Ranked STAPLE for template based segmentation propagation. MICCAI workshop on Multi-Atlas Labeling and Statistical Fusion.

Cardoso, M. J., Melbourne, A., Kendall, G. S., Modat, M., Hagmann, C. F., Robertson, N. J., Marlow, N., Ourselin, S., 2011d. Adaptive Neonatal Brain Segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Lecture Notes in Computer Science, pp. 378–386.

Clarkson, M. J., Cardoso, M. J., Ridgway, G. R., Modat, M., Leung, K. K., Rohrer, J. D., Fox, N. C., Ourselin, S., May 2011. A comparison of voxel and surface based cortical thickness estimation methods. NeuroImage.

Cointepas, Y., Bloch, I., Garnero, L., 2001. A cellular model for multi-objects multi-dimensional homotopic deformations. Pattern Recognition 34 (9), 1785–1798.

Collins, D. L., Pruessner, J. C., Oct. 2010. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. NeuroImage 52 (4), 1355–1366.

Crum, W. R., Camara, O., Hill, D. L. G., 2006. Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. IEEE Transactions on Medical Imaging 25 (11), 1451–1461.

Das, S. R., Avants, B. B., Grossman, M., Gee, J. C., Apr. 2009. Registration based cortical thickness measurement. NeuroImage 45 (3), 867–879.

Desikan, R. S. R., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., Schmansky, N. J., Greve, D. N., Salat, D. H., Buckner, R. L., Fischl, B., Alzheimer's Disease Neuroimaging Initiative, Aug. 2009. Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 132 (Pt 8), 2048–2057.

Desikan, R. S. R., Fischl, B., Cabral, H. J., Kemper, T. L., Guttmann, C. R. G., Blacker, D., Hyman, B. T., Albert, M. S., Killiany, R. J., Sep. 2008. MRI measures of temporoparietal regions show differential rates of atrophy during prodromal AD. Neurology 71 (11), 819–825.

Du, A.-T., Schuff, N., Kramer, J. H., Rosen, H. J., Gorno-Tempini, M. L., Rankin, K., Miller, B. L., Weiner, M. W., Apr. 2007. Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia. Brain 130 (Pt 4), 1159–1166.

Dubois, J., Benders, M., Borradori-Tolsa, C., Cachia, A., Lazeyras, F., Leuchter, R. H.-V., Sizonenko, S. V., Warfield, S. K., Mangin, J.-F. F., Huppi, P. S., Aug. 2008. Primary cortical folding in the human newborn: an early marker of later functional development. Brain 131 (Pt 8), 2028–2041.

Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A., Walhovd, K. B., Oct. 2010. Effects of memory training on cortical thickness in the elderly. NeuroImage 52 (4), 1667–1676.

Evans, A. C., Collins, D. L., Mills, S., Brown, E., Kelly, R., Peters, T., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference 3, 1813–1817.

Fischl, B., Dale, A. M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the National Academy of Sciences of the United States of America 97 (20), 11050–11055.

Fischl, B., Salat, D. H., Busa, E., Albert, M. S., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R. J., Kennedy, D. N., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A. M., 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Gelman, N., Gorell, J., Barker, P., Savage, R. M., Spickler, E. M., Windham, J. P., Knight, R. A., Jan. 1999. MR imaging of human brain at 3.0 T: preliminary report on transverse relaxation rates and relation to estimated iron content. Radiology 1 (210), 759–767.

Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R., 2009. On the manifold structure of the space of brain images. Medical Image Computing and Computer-Assisted Intervention 12, 305–312.

Gilmore, J. H., Lin, W., Prastawa, M. W., Looney, C. B., Vetsa, Y. S. K., Knickmeyer, R. C., Evans, D. D., Smith, J. K., Hamer, R. M., Lieberman, J. A., Gerig, G., Feb. 2007. Regional gray matter

growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. J Neurosci 27 (6), 1255–1260.

Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., Mitchell, T. N., Brooks, D. J., Duncan, J. S., Aug. 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Human Brain Mapping 19 (4), 224–247.

Hammers, A., Chen, C.-H., Lemieux, L., Allom, R., Vossos, S., Free, S. L., Myers, R., Brooks, D. J., Duncan, J. S., Koepp, M. J., Jan. 2007. Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space. Human Brain Mapping 28 (1), 34–48.

Han, L., Hipwell, J. H., Tanner, C., Taylor, Z. A., Mertzanidou, T., Cardoso, M. J., Ourselin, S., Hawkes, D. J., Jan. 2012. Development of patient-specific biomechanical models for predicting large breast deformation. Physics in Medicine and Biology 57 (2), 455–472.

Han, X., Pham, D. L., Tosun, D., Rettmann, M. E., Xu, C., Prince, J. L., Nov. 2004. CRUISE: cortical reconstruction using implicit surface evolution. NeuroImage 23 (3), 997–1012.

Heckemann, R. A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J. V., Hammers, A., Alzheimer's Disease Neuroimaging Initiative, May 2010. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. NeuroImage 51 (1), 221–227.

Holland, D., Brewer, J. B., Hagler, D. J., Fenema-Notestine, C., Fenema-Notestine, C., Dale, A. M., Alzheimer's Disease Neuroimaging Initiative, Dec. 2009. Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proceedings of the National Academy of Sciences of the United States of America 106 (49), 20954–20959.

Hutton, C., De Vita, E., Ashburner, J., Deichmann, R., Turner, R., 2008. Voxel-based cortical thickness measurements in MRI. NeuroImage 40 (4), 1701–1710.

Ince, D. C., Hatton, L., Graham-Cumming, J., Feb. 2012. The case for open computer programs. Nature 482 (7386), 485–488.

Jernigan, T. L., Tallal, P., May 1990. Late childhood changes in brain morphology observable with MRI. Developmental medicine and child neurology 32 (5), 379–385.

Jones, S. E., Buchbinder, B. R., Aharon, I., 2000. Three-dimensional mapping of cortical thickness using Laplace's equation. Human Brain Mapping 11 (1), 12–32.

Kapellou, O., Counsell, S. J., Kennea, N., Dyet, L. E., Saeed, N., Stark, J., Maalouf, E. F., Duggan, P. J., Ajayi-Obe, M., Hajnal, J. V., Allsop, J. M., Boardman, J. P., Rutherford, M. A., Cowan, F. M., Edwards, A. D., Aug. 2006. Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. PLoS Med 3 (8), e265.

Khalimsky, E. D., Kopperman, R., Meyer, P. R., Jul. 1990. Computer graphics and connected topologies on finite ordered sets. Topology and its Applications 36 (1).

Khwaja, O., Volpe, J. J., Mar. 2008. Pathogenesis of cerebral white matter injury of prematurity. Arch Dis Child Fetal Neonatal Ed 93 (2), F153–F161.

Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J. M., Kim, S. I., Evans, A. C., Aug. 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. NeuroImage 27 (1), 210–221.

Kitamoto, A., Takagi, M., 1999. Image Classification Using Probabilistic Models that Reflect the Internal Structure of Mixels. Pattern Analysis and Applications 2.

Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., Apr. 2009. Evaluation of automated brain MR image segmentation and volumetry methods. Human Brain Mapping 30 (4), 1310–1327.

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D. D., Wilber, K., Smith, J. K., Hamer, R. M., Lin, W., Gerig, G., Gilmore, J. H., Nov. 2008. A structural MRI study of human brain development from birth to 2 years. J Neurosci 28 (47), 12176–12182.

Kondor, R. I., Lafferty, J., Jan. 2002. Diffusion kernels on graphs and other discrete input spaces. ICML.

Kovalevsky, V. A., 1989. Finite Topology as Applied to Image-Analysis. Computer Vision Graphics and Image Processing 46 (2), 141–161.

Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S. J., Gousias, I. S., Boardman, J. P., Rutherford, M. A., Edwards, A. D., Hajnal, J. V., Rueckert, D., 2011. A dynamic 4D probabilistic atlas of the developing brain. NeuroImage 54 (4).

Lam, L., Suen, C. Y., Sep. 1995. Optimal combinations of pattern classifiers. Pattern Recognition Letters 16 (9), 945–954.

Lehmann, M., Crutch, S. J., Ridgway, G. R., Ridha, B. H., Barnes, J., Warrington, E. K., Rossor, M. N., Fox, N. C., Sep. 2009. Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. Neurobiology of Aging.

Lerch, J., Pruessner, J. C., Zijdenbos, A. P., Hampel, H., Teipel, S., Evans, A. C., Jul. 2005. Focal Decline of Cortical Thickness in Alzheimer's Disease Identified by Computational Neuroanatomy. Cerebral Cortex 15 (7), 995.

Leung, K. K., Barnes, J., Ridgway, G. R., Bartlett, J. W., Clarkson, M. J., Macdonald, K., Schuff, N., Fox, N. C., Ourselin, S., Alzheimer's Disease Neuroimaging Initiative, Jul. 2010. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. NeuroImage 51 (4), 1345–1359.

Lohmann, G., Preul, C., Hund-Georgiadis, M., 2003. Morphology-based cortical thickness estimation. In: Information Processing for Medical Imaging. pp. 89–100.

Lötjönen, J., Wolz, R., Koikkalainen, J., Julkunen, V., Thurfjell, L., Lundqvist, R., Waldemar, G., Soininen, H., Rueckert, D., Alzheimer's Disease Neuroimaging Initiative, May 2011. Fast and robust extraction of hippocampus from MR images for diagnostics of Alzheimer's disease. NeuroImage 56 (1), 185–196.

MacDonald, D., Kabani, N., Avis, D., Evans, A. C., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. NeuroImage 12 (3), 340.

Mangham, L. J., Petrou, S., Doyle, L. W., Draper, E. S., Marlow, N., Feb. 2009. The cost of preterm birth throughout childhood in England and Wales. Pediatrics 123 (2), e312–e327.

Mangin, J.-F. F., Frouin, V., Bloch, I., Régis, J., 1995. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. Journal of Mathematical Imaging and Vision 5 (4), 297–318.

Marlow, N., Wolke, D., Bracewell, M. A. A., Samara, M., 2005. Neurologic and Developmental Disability at Six Years of Age after Extremely Preterm Birth. New England Journal of Medicine 352 (1), 9–19.

Mathur, A., Inder, T. E., 2009. Magnetic resonance imaging–insights into brain injury and outcomes in premature infants. J Commun Disord 42 (4), 248–255.

Ment, L. R., Hirtz, D., Hüppi, P. S., Nov. 2009. Imaging biomarkers of outcome in the developing preterm brain. Lancet Neurol 8 (11), 1042–1055.

Mertzanidou, T., Hipwell, J. H., Cardoso, M. J., Zhang, X., Tanner, C., Ourselin, S., Bick, U., Huisman, H., Karssemeijer, N., Hawkes, D. J., Mar. 2012. MRI to X-ray mammography registration using a volume-preserving affine transformation. Medical Image Analysis.

Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., Ourselin, S., Jun. 2010. Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine 98 (3), 278–284.

Morris, R. D., Descombes, X., Zerubia, J., Sep. 1996. The Ising/Potts model is not well suited to segmentation tasks. IEEE Digital Signal Processing Workshop.

Nesvåg, R., Lawyer, G., Varnäs, K., Fjell, A. M., Walhovd, K. B., Frigessi, A., Jönsson, E. G., Agartz, I., 2008. Regional thinning of the cerebral cortex in schizophrenia: effects of diagnosis, age and antipsychotic medication. Schizophrenia research 98 (1-3), 16–28.

Oishi, K., Mori, S., Donohue, P. K., Ernst, T., Anderson, L., Buchthal, S., Faria, A., Jiang, H., Li, X., Miller, M. I., van Zijl, P. C. M., Chang, L., May 2011. Multi-contrast human neonatal brain atlas: application to normal neonate development analysis. NeuroImage 56 (1), 8–20.

Ourselin, S., Roche, A., Prima, S., Ayache, N., 2000. Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In: Goos, G., Hartmanis, J., Leeuwen, J., Delp, S. L., DiGoia, A. M., Jaramaz, B. (Eds.), Medical Image Computing and Computer-Assisted Intervention. MICCAI 2000, pp. 557–566.

Ourselin, S., Roche, A., Subsol, G., Pennec, X., 2001. Reconstructing a 3D structure from serial histological sections. Image and Vision Computing 19, 25–31.

Passat, N., Couprie, M., Bertrand, G., Nov. 2008. Minimal Simple Pairs in the 3-D Cubic Grid. Journal of Mathematical Imaging and Vision 32 (3), 239–249.

Patenaude, B., Smith, S. M., Kennedy, D. N., Jenkinson, M., Jun. 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage 56 (3), 907–922.

Pham, D. L., 2002a. Fuzzy clustering with spatial constraints. In: International Conference on Image Processing. IEEE, pp. II–65–II–68 vol.2.

Pham, D. L., Jan. 2002b. Robust fuzzy segmentation of magnetic resonance images. Computer-Based Medical Systems, 127–131.

Pham, D. L., Prince, J. L., Sep. 1999. Adaptive fuzzy segmentation of magnetic resonance images. IEEE Transactions on Medical Imaging 18 (9), 737–752.

Prastawa, M. W., Gilmore, J. H., Lin, W., Gerig, G., Oct. 2005. Automatic segmentation of MR images of the developing newborn brain. Med Image Anal 9 (5), 457–466.

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.-A., Démonet, J.-F., Duret, V., Puel, M., Berry, I., Fort, J.-C., Celsis, P., The Alzheimer's Disease Neuroimaging Initiative, 2009. Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain 8 (132), 2036–2047.

Rex, D. E., Shattuck, D. W., Woods, R. P., Narr, K. L., Luders, E., Rehm, K., Stoltzner, S., Rottenberg, D. A., Toga, A. W., Oct. 2004. A meta-algorithm for brain extraction in MRI. NeuroImage 23, 625–637.

Rocha, K. R., Yezzi Jr, A. J., Prince, J. L., 2005. A hybrid Eulerian-Lagrangian approach for thickness, correspondence, and gridding of annular tissues. IEEE Transactions on Image Processing 3765, 72–81.

Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr, C. R., Apr. 2004a. Evaluation of Atlas Selection Strategies for Atlas-Based Image Segmentation with Application to Confocal Microscopy Images of Bee Brains. NeuroImage 21 (4), 1428–1442.

Rohlfing, T., Russakoff, D. B., Maurer Jr, C. R., Aug. 2004b. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Transactions on Medical Imaging 23 (8), 983–994.

Rosas, D., Salat, D. H., Lee, S. Y., Zaleta, A. K., Pappu, V., Fischl, B., Greve, D. N., Hevelone, N., Hersch, S. M., Apr. 2008. Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. Brain 131 (Pt 4), 1057–1068.

Roth, H. R., McClelland, J. R., Boone, D. J., Modat, M., Cardoso, M. J., Hampshire, T. E., Hu, M., Punwani, S., Ourselin, S., Slabaugh, G. G., Halligan, S., Hawkes, D. J., Jun. 2011. Registration of the endoluminal surfaces of the colon derived from prone and supine CT colonography. Medical physics 38 (6), 3077–3089.

Ruan, S., Jaggi, C., Fadili, J., Bloyet, D., Dec. 2000. Brain Tissue Classification of Magnetic Resonance Images Using Partial Volume Modeling. IEEE Transactions on Medical Imaging 19 (12).

Rutherford, M. A., Supramaniam, V., Ederies, A., Chew, A. T. M., Bassi, L., Groppo, M., Anjari, M., Counsell, S. J., Ramenghi, L. A., Jun. 2010. Magnetic resonance imaging of white matter diseases of prematurity. Neuroradiology 52 (6), 505–521.

Sabuncu, M. R., Desikan, R. S. R., Sepulcre, J., Yeo, B. T. T., Liu, H., Schmansky, N. J., Reuter, M., Weiner, M. W., Buckner, R. L., Sperling, R. A., Fischl, B., Alzheimer's Disease Neuroimaging Initiative, Aug. 2011. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. Archives of Neurology 68 (8), 1040–1048.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., Morris, J. C., Dale, A. M., Fischl, B., Jul. 2004. Thinning of the cerebral cortex in aging. Cerebral Cortex 14 (7), 721–730.

Scott, M. L. J., Bromiley, P. A., Thacker, N., Hutchinson, C. E., Jackson, A., Apr. 2009. A fast, model-independent method for cerebral cortical thickness estimation using MRI. Medical Image Analysis 13 (2), 269–285.

Seghier, M. L., Ramlackhansingh, A., Crinion, J., Leff, A. P., Price, C. J., 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. NeuroImage 41 (4), 1253–1266.

Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D. H., Hahn, H., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. NeuroImage 22 (3), 1060–1075.

Shefer, V. F., 1973. Absolute number of neurons and thickness of the cerebral cortex during aging, senile and vascular dementia, and Pick's and Alzheimer's diseases. Neuroscience and Behavioral Physiology 6 (4), 319–324.

Shiee, N., Bazin, P.-L., Cuzzocreo, J. L., Blitz, A., Pham, D. L., 2011. Segmentation of brain images using adaptive atlases with application to ventriculomegaly. Information Processing for Medical Imaging 22, 1–12.

Smith, S. M., 2002. Fast robust automated brain extraction. Human Brain Mapping 17 (3), 143–155.

Song, Z., Awate, S. P., Licht, D. J., Gee, J. C., 2007. Clinical Neonatal Brain MRI Segmentation Using Adaptive Nonparametric Data Models. In: Ayache, N., Ourselin, S., Maeder, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention. Departments of Radiology, University of Pennsylvania, Springer Berlin / Heidelberg, pp. 883–890.

Souvenir, R., Pless, R., Mar. 2007. Image distance functions for manifold learning. Image and Vision Computing 25 (3), 365–373.

Tang, H., Wu, E., Ma, Q., Gallagher, D., Perera, G., Zhuang, T., 2000. MRI brain image segmentation by multi-resolution edge detection and region selection. Computerized Medical Imaging and Graphics 24 (6), 349–357.

Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J. L., 2010. Longitudinal Changes in Cortical Thickness Associated with Normal Aging. NeuroImage.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15 (1), 273–289.

Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P., 2001. Automated segmentation of multiple sclerosis lesions by modeloutlier detection. IEEE Transactions on Medical Imaging 20 (8), 677–688.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999a. Automated model-based bias field correction of MR images of the brain. IEEE Transactions on Medical Imaging 18 (10), 885–896.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999b. Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging 18 (10), 897–908.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 2003. A unifying framework for partial volume segmentation of brain MR images. IEEE Transactions on Medical Imaging 22 (1), 105–119.

Volpe, J. J., Jan. 2009. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. Lancet Neurol 8 (1), 110–124.

Wang, H., Fei, B., 2009. A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme. Medical Image Analysis 13 (2), 193–202.

Warfield, S. K., Zou, K. H., Wells III, W. M., Jul. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23 (7), 903–921.

Weisenfeld, N. I., Warfield, S. K., Aug. 2009. Automatic segmentation of newborn brain MRI. NeuroImage 47 (2), 564–572.

Wells III, W. M., Grimson, W. E. L., Kikinis, R., Jolesz, F. A., 1996. Adaptive segmentation of MRI data. IEEE Transactions on Medical Imaging 15 (4), 429–442.

Wolz, R., Aljabar, P., Hajnal, J. V., Hammers, A., Rueckert, D., Alzheimer's Disease Neuroimaging Initiative, Jan. 2010. LEAP: learning embeddings for atlas propagation. NeuroImage 49 (2), 1316–1325.

Woods, K., Kegelmeyer, W. P., Bowyer, K., Apr. 1997. Combination of multiple classifiers using local accuracy estimates. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (4), 405–410.

Xu, L., Krzyzak, A., Suen, C. Y., 1992. Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Transactions Syst., Man, Cybern. 22 (3), 418–435.

Xue, H., Srinivasan, L., Jiang, S., Rutherford, M. A., Edwards, A. D., Rueckert, D., Hajnal, J. V., Nov. 2007. Automatic segmentation and reconstruction of the cortex from neonatal MRI. NeuroImage 38 (3), 461–477.

Xue, Z., Shen, D., Davatzikos, C., Apr. 2006. CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. NeuroImage 30 (2), 388–399.

Yeo, B. T. T., Sabuncu, M. R., Desikan, R. S. R., Fischl, B., Golland, P., Oct. 2008. Effects of registration regularization and atlas sharpness on segmentation accuracy. Medical Image Analysis 12 (5), 603–615.

Yezzi Jr, A. J., Prince, J. L., 2003. An Eulerian PDE approach for computing tissue thickness. IEEE Transactions on Medical Imaging 22 (10), 1332–1339.

Yu, X., Zhang, Y., Lasky, R. E., Parikh, N. A., Narayana, P. A., 2010. Comprehensive Brain MRI Segmentation in High Risk Preterm Newborns. PLoS ONE 5 (11).

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R., Ho, S., Gee, J. C., Gerig, G., Jul. 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 31 (3), 1116–1128.

Yushkevich, P. A., Wang, H., Pluta, J., Das, S. R., Craige, C., Avants, B. B., Weiner, M. W., Mueller, S., Dec. 2010. Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. NeuroImage 53 (4), 1208–1224.

Zhang, J., 1992. The mean field theory in EM procedures for Markov random fields. IEEE Transactions on Signal Processing 40 (10), 2570–2583.

Zhang, Y., Brady, M., Smith, S. M., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging 20 (1), 45–57.

Zijdenbos, A. P., Dawant, B. M., Margolin, R. A., Palmer, A. C., 1994. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Transactions on Medical Imaging 13 (4), 716–724.